Forensic sampling and DNA databases: background/issues paper

Commissioned by:
Victorian Parliament Law Reform Committee

Written by:
Dr Peta Stringer (Manager, DNA Science Branch)
Victoria Forensic Science Centre
Forensic Drive
Macleod 3085
Victoria
Tel: 03 94503444
Fax: 03 94503601
Email: peta.stringer@police.vic.gov.au

Section 2.1.2 (with the exception of Victorian DNA Database match information) Written by:
John Anderson, Crimtrac DNA Project Manager and Jonathan D Mobbs, CEO, The Crimtrac Agency, Canberra

Date:
June 2002
Index:

VICTORIAN PARLIAMENT LAW REFORM COMMITTEE..................................................... 1
BRIEF: ......................................................................................................................... 4
TERMS OF REFERENCE: ............................................................................................. 4
DISCLAIMER: .............................................................................................................. 4

1 THE COLLECTION, PROCESSING AND STORAGE OF DNA SAMPLES ..................... 5
  1.1 The Collection of DNA samples .............................................................................. 5
    1.1.1 Reference samples ............................................................................................ 5
    1.1.2 Crime scene samples ........................................................................................ 6
  1.2 Processing of DNA samples .................................................................................... 10
    1.2.1 DNA extraction .................................................................................................. 12
      1.2.1.1 Non-organic extraction methods ................................................................. 12
      1.2.1.1.1 Chelex ....................................................................................................... 12
      1.2.1.1.2 Magnetic beads ......................................................................................... 13
      1.2.1.1.3 FTA® paper ............................................................................................... 13
      1.2.1.3 Further purification of extracted DNA ....................................................... 15
    1.2.2 DNA Quantitation .............................................................................................. 15
      1.2.2.1 Quantiblot .................................................................................................... 15
    1.2.3 Amplification of DNA ....................................................................................... 17
    1.2.4 Fragment Separation ........................................................................................ 20
    1.2.5 Quality Management and Accreditation ......................................................... 24
  1.3 Storage of DNA samples .......................................................................................... 25
  2 THE ESTABLISHMENT AND USE OF A DNA DATABASE ..................................... 26
  2.1 The establishment of DNA databases ..................................................................... 26
    2.1.1 General .............................................................................................................. 26
    2.1.2 Australia .......................................................................................................... 26
      2.1.2.1 The establishment of the Australian National Criminal Investigation DNA Database (NCIDD) ............................................................................................... 27
      2.1.2.2 Interpol ........................................................................................................ 37
  2.2 The use of a DNA database .................................................................................... 38
    2.2.1 DNA database searches and the meaning of matches obtained ....................... 38
    2.2.2 What profiles should be put on a DNA database-single source DNA samples and/or DNA mixture profiles? ............................................................. 42
    2.2.3 Frequency of DNA database match runs ....................................................... 44
    2.2.4 Permissible matching of DNA profiles ........................................................... 45
    2.2.5 Destruction of information .............................................................................. 45
  3 REFERENCES ............................................................................................................ 47
  4 GLOSSARY .............................................................................................................. 52

List of figures:

FIGURE 1: A DIAGRAMMATIC REPRESENTATION OF THE QUANTIBLOT PROCESS ............................................................... 16
FIGURE 2: PCR PROCESS ..................................................................................................... 20
FIGURE 3: ELECTROPHOREGRAM SHOWING THE GENOTYPER PLOT OF A SAMPLE CONTAINING DNA FROM A SINGLE SOURCE .................................................................................... 22
FIGURE 4: ELECTROPHOREGRAM SHOWING THE GENOTYPER PLOT FOR A SAMPLE CONTAINING A MIXTURE OF DNA FROM MORE THAN ONE SOURCE ................................................... 23
FIGURE 5: NCIDD NETWORK ......................................................................................... 30
FIGURE 6: GOVERNANCE MODEL FOR NCIDD ........................................................................ 32
List of tables:

TABLE 1: AMPFLSTR PROFILER PLUS™ PCR AMPLIFICATION KIT STRS ANALYSED..........................18
TABLE 2: AMPLIFICATION PARAMETERS FOR AMPFLSTR PROFILER PLUS™ AMPLIFICATION REACTIONS .................................................................................................................19
TABLE 3: MATRIX FILE TABLE EXAMPLE.........................................................................................................................21
SOURCE: CHIEF CONSTABLE BEN GUNN, CAMBRIDGESHIRE CONSTABULARY, UK, 2002
TABLE 4: UK NDNADB FIGURES .......................................................................................................................34
TABLE 5: VICTORIAN DNA DATABASE MATCHES DETECTED...........................................................................35
TABLE 6: BREAKDOWN OF VICTORIAN DNA DATABASE MATCHES BY OFFENCE TYPE........................................36
TABLE 7: DNA PROFILE RESULTS...........................................................................................................40
TABLE 8: LIKELIHOOD RATIOS FOR MATCHES BETWEEN PROFILES ..........................................................41
TABLE 9: PERMISSIBLE MATCHING OF DNA PROFILES UNDER §464ZGI OF THE CRIMES (DNA DATABASE) ACT 2002 .................................................................45
Brief:
To produce a background /issues paper for the Victorian Parliament Law Reform Committee in relation to its inquiry into Forensic Sampling and DNA Databases. The terms of reference for the inquiry follow.

The paper will identify and discuss the practical and technical considerations relating to:
- The collection, processing and storage of DNA samples; and
- The establishment and use of a DNA database.

Terms of reference:
That, pursuant to the Parliamentary Committees Act 1968, The Law Reform Committee be required to inquire into, consider and report on the following:

“The collection, use and effectiveness of forensic sampling and the use of DNA databases in criminal investigations, with particular emphasis on identifying areas and procedures which would more effectively utilise forensic sampling and improve investigation and detection of crime.”

Disclaimer:
Whilst some scientific products and suppliers are mentioned in this document to illustrate certain points, this does not necessarily infer that the Victoria Forensic Science Centre prefers these products over other products, or that other scientific products and suppliers are not equally as reliable and suitable.
1 The collection, processing and storage of DNA samples

1.1 The collection of DNA samples
DNA samples broadly fall into two main categories: ‘reference’ samples from individuals, and crime scene samples. Samples are collected from individuals to determine whether they are potential donors to DNA detected in crime scene samples or not.

The discrimination power of DNA testing is such that individuals can frequently be conclusively eliminated as potential participants in a criminal offence. This is clearly beneficial in terms of the progress of a criminal investigation.

Where an individual cannot be excluded, DNA evidence is typically able to provide quality evidence to this effect.

1.1.1 Reference samples
Reference samples can consist of any body fluid or tissue from a person. Reference samples are used to determine the DNA profile type of the person allowing comparison of that profile with any DNA profile detected in other samples. Generally, reference samples contain sufficient DNA of good quality for DNA profiling. The same may not be true of crime scene samples.

The most commonly used reference samples include blood, saliva, and hair. This is due to the fact that blood and saliva in particular contain abundant, easily accessible DNA. Hair is another readily available source of DNA that can be easily obtained from the body, although the amount of genomic DNA in a single hair is dependent upon the growth phase of the hair. Muscle, bone, and teeth can also be used as reference samples. These sample types are often typically taken from deceased persons where blood, saliva or hair may not be available or where those sources are degraded.

Forensic laboratories prefer that reference samples be taken using particular substrates for collection. Through laboratory validation studies, it is often shown that some substrates are more reliable than others in terms of their ability to preserve biological material for successful forensic testing.

It is also preferable that samples from different individuals are collected using the same type of substrate. Typically, this is usually to enable batch processing of samples in the laboratory using only one DNA testing methodology. However, it is also the preferred approach, as it means that fewer resources are required for training police and medical personnel in the taking of samples. In addition, reduced resources are required for training about specific requirements that may be needed to preserve sample integrity. This includes legal issues around sample collection and the prevention of contamination of samples with DNA other than the sample donor’s DNA.

Some of the substrate types that are currently available for the collection of blood and buccal (saliva) samples are swabs, isocode stix (Medos), FTA paper (Crown
Scientific). There are other substrates that can be used. Local requirements, experiences and preferences often determine the choice of substrate.

### 1.1.2 Crime scene samples

Crime scene samples can be from a person (e.g., vaginal swabs in the case of sexual assault), from items found at a crime scene, or from items associated with the commission of a crime which may be located away from the actual crime scene itself.

Biological crime samples can potentially contain large or small amounts of biological material, cover small or extensive areas, be degraded or non-degraded, contain material from one or multiple persons, contain material that may have been recently deposited or material that may have been present for many years, and be visible or not visible to the naked eye. All these variables mean that there is a multiplicity of crime sample types. The types of surfaces that these samples may be deposited on may be extremely variable, and can differ widely depending on whether the sample has been deposited indoors or outdoors.

In many instances before samples can be collected at crime scenes it is necessary to locate them. In the case of visible bloodstaining this can be obvious. However, not all body fluid stains are visible to the naked eye, and some may be difficult to locate depending on the surface they are deposited on.

Whilst some specific tests are referred to in subsequent paragraphs of this report, there are a wide variety of other screening tests that can be used to locate various body fluid stains. Many of these tests have undergone substantial modification for forensic use since their first inception.

Visible blood can be initially detected by the naked eye, and there are a number of simple chemical tests that will allow the analyst to confirm that blood may be present. Examples include the o-toluidine (1, 2) and haemochromogen (Takayama) tests (3,4). In addition, there are now commercially available ‘test’ or ‘reagent strips’ that can be used to detect the possible presence of blood, for example Bayer Mulitiple® Reagent Strips for Urinalysis. These strips are commonly called ‘haemastix’ and can indicate the presence of blood through a colour change reaction on the strip. Confirmation of apparent blood detected by haemastix, for example by the haemochromogen test, would still be required to confirm the presence of blood.

It also must be noted that whilst some of these tests may confirm the presence of blood, additional testing of other types must be performed to confirm the blood as being human in origin, for example Ouchterlony double immuno-diffusion using anti-human haemoglobin antisera (5,6).

Blood that has been diluted and is therefore not visible to the naked eye can be visualised using a test called the luminol test (7-12). Diluted blood can be a result of deliberate attempts to remove visible bloodstaining, for example, washing clothing or surfaces. Blood can also be diluted when an exterior crime scene is subjected to rain.

The luminol test detects a breakdown product of blood. The older and more degraded the blood, the more likely it is that the test will detect the presence of blood.
The luminol test involves spraying areas or surfaces suspected of being bloodstained with a chemical solution. The test must be performed in the dark. When blood, which may have been diluted many times, has been detected, a blue-ish coloured luminescence or glow can be observed. This test is also not specific for blood. Any substance which has peroxidase activity or 'peroxidase-like' activity (eg some plants, some varnishes) can induce a similar luminescence. Typically however, the luminescence produced by non-blood sources is of a different colour, usually more white than blue, although it may be difficult for an inexperienced user of the test to distinguish the difference.

Once a luminol positive area has been identified, a sample of the surface can be removed or swabbed for further laboratory testing to attempt to confirm the presence of blood, and the species origin of the blood, if possible. However, confirmation of the presence of blood and/or determination of it’s species origin can be difficult as the process of luminol testing further dilutes the already diluted blood that may have been present and contaminates it with other substances. In addition, it can also be very difficult to obtain a DNA profile from the trace amounts of DNA present for the same reasons.

It is frequently necessary to locate semen staining in cases of sexual assaults. This can be done by using a chemical test to detect the presence of the protein acid phosphatase (AP) (13-15). This is a protein and enzyme that is found in both vaginal fluid and semen. Typically males have considerably elevated levels of acid phosphatase over females, however, at the extremes there are males with low AP levels, and females with high AP levels. For this reason, this test is also not specific for the identification of semen. In addition, there are many other animals that also have AP in their body fluids, and other substances (eg some plant materials) that can induce a similar positive result to this test as that induced by semen and vaginal fluid (15-18). For this reason, this test can only ever be indicative of the presence of semen. Additionally, it is very likely that a semen stain will also contain vaginal fluid, for example on swabs from the female genital tract or from underwear worn by women.

The AP test is performed by pressing wet white paper onto a surface suspected of containing semen and/or vaginal fluid. Since AP is a water-soluble protein it is readily absorbed into the paper. Once this has been done the paper is removed and sprayed with a chemical solution. Where the chemical solution comes into contact with the AP, a colour change from colourless to purple ensues. By re-positioning the paper over the exhibit, the location of the AP positive area can be determined.

To confirm the presence of semen additional testing in the laboratory must be performed to identify the presence of spermatozoa, or to look for the presence of other proteins/substances unique to semen (eg prostrate specific antigen-PSA) (19,20).

Spermatozoa are cells that are present in semen and that are unique to semen. Human spermatozoa have a characteristic morphology (ie shape), and can be identified by an experienced analyst. Identification of their presence confirms the presence of semen. Staining cells on a microscope slide that has been prepared from the exhibit can identify these cells. There are a number of staining techniques currently available (eg haemotoxylin/eosin (21,22), Christmas Tree staining (23,24).
Vaginal fluid may be located by AP screening if a female has high levels of this protein in her body. Further testing to indicate the presence of epithelial cells (e.g., Lugols iodine (25,26)) can be done in the laboratory. However, confirmation of the presence of vaginal fluid in forensic samples can be very difficult, since as detailed previously, there are other body fluids that also contain acid phosphatase.

Faeces can be partially identified by its smell and appearance, however, a test to detect urobilinogen (a component of faeces and a breakdown product of bile) (27,28) can be performed in the laboratory. It may or may not be possible to determine that the faeces is human in origin, or of animal origin. The determination that faeces is human in origin can be by Ouchterlony double immuno-diffusion testing where the detection of anti-human antibodies is attempted. This may or may not be successful depending on how much human cellular material is present in the faeces.

Saliva can be detected by performing a test to detect the presence of α-amylase. This is a protein present in saliva. This protein can be detected via a screening test using ‘pink paper’ (29). White paper is pre-treated with a pink coloured dye. The α-amylase on the surface of an exhibit is readily absorbed into the paper when the paper is wet and has been pressed onto the surface suspected to contain amylase. It is possible then to locate areas on the paper where α-amylase may be present as the pink dye is converted to a colourless form and white areas appear on the paper. By re-positioning the paper back over the exhibit it is possible to locate the area on the exhibit that potentially had α-amylase activity.

Urine can be suspected to be present on an exhibit because of its smell. To confirm the presence of urine it is necessary to conduct a urine test. This can be done using a commercially available testing kit, the Merckognost Urea Rapid Screening Test Kit (Diagnostica Merck, Germany).

Aside from these main body fluid types, there are few screening tests suitable for forensic samples available to identify other body fluid types or tissues. For most other suspected body fluid types or tissues, samples are taken from a crime scene and tested in the laboratory. This testing is usually unable to involve definite identification of the body fluid or tissue type. The approach is to perform DNA testing on the suspected body fluid or tissue, having selected a general testing procedure that will most likely result in a DNA profile irrespective of the source of the biological material.

It is becoming increasingly more difficult to identify the specific biological material that may be present on an exhibit. This is due to the ongoing development of more sensitive sample collection methods and DNA profiling techniques. Traditionally, identification of biological material could only be performed on material that was visible to the naked eye, due to the limited sensitivity of testing methods available at the time. Now, it is possible to swab areas of exhibits that are only suspected to contain biological material, and from this obtain a DNA profile, and this capability has significantly impacted on the degree in which forensic testing may assist crime investigation. It also means that DNA profiles can be obtained from biological material where the actual biological origin of that biological material cannot be determined.
In the case of suspected bone, a particular DNA testing procedure can be followed. This is based on the method of Hagelburg (30). Bone, being a hard substance and containing limited amounts of DNA relative to other body fluids and tissues, can be difficult to extract DNA from, and requires a more severe DNA extraction process than other tissue types.

To assist a criminal investigation it is often of relevance to be able to confirm that a biological sample was human in origin as opposed to having originated from another animal species. A test to determine species origin (Ouchterlony double immuno-diffusion testing) can be performed where there is a relatively large amount of sample present. However, sometimes there is insufficient biological material present to enable this test to be performed. In that case it may be possible to infer that the biological material being tested is of higher primate origin, ie either from humans or gorillas. This inference can be made simply because the DNA extracted from the sample produces a result in a test to quantitate the amount of extracted DNA obtained during the DNA profiling process. This particular quantitation test (the Quantiblot test, refer to section 1.2.2.1) will only work for higher primate DNA. Therefore if the sample cannot directly be shown to be human in origin either by morphological identification (eg for spermatozoa) or by Ouchterlony double immuno-diffusion testing, it may be identified as being of higher primate origin via the DNA quantitation process. The actual steps involved in DNA profiling will be examined in greater detail in section 1.2 of this report.

Samples at crime scenes can be collected in a variety of ways in terms of collection for DNA analyses. The simplest way is to collect the entire exhibit, and package it in a container that can be sealed and labelled.

Paper bags are often the most suitable form of packaging. Plastic bags are generally a poor packaging option as they encourage bacterial/fungal growth on exhibits that may be damp. This can ultimately render exhibits unsuitable for forensic analysis. Exhibits should be packaged such that their transport to the laboratory does not compromise forensic testing or the condition of the exhibit itself.

Adequate sealing of exhibit packaging is to ensure that the exhibit is not exposed to other people, and therefore is not altered from the state in which it was collected prior to its analysis at a forensic laboratory. The labelling needs to be sufficient to identify the exhibit and distinguish it from any other exhibit that may be collected in relation to the same case or any other case. Collecting exhibits in this manner means that the forensic scientist can then take appropriate samples from the exhibit for DNA analyses, and perform additional other (non-DNA) analyses as may be required.

It must be remembered that DNA profiling results will only be as good as the potential source from which they were acquired. Poor treatment of exhibits leading either to their degradation or destruction, and inadequate labelling of exhibits has the potential to minimise the value of any forensic testing that may be applied to them.

If an exhibit cannot be delivered to a forensic laboratory either because it is too large or cumbersome, on a surface from which removal is difficult (eg road surface), where the sample could be lost through packaging the exhibit (eg a hair, or blood flakes), or
where the rest of the surface or exhibit is not of relevance, a sample may be removed for laboratory analysis.

Samples can be removed by simply picking the sample off the object. If this is done it is essential that the person collecting the sample collects it in a manner whereby they do not add their own DNA to the sample, or anyone else's. Making sure that the person wears adequate protective clothing, plastic/latex gloves, and a facemask can facilitate this.

Although forensic scientists are used to examining exhibits in this manner, this level of use of protective clothing is not commonplace amongst investigators. Given that the sensitivity of forensic tests is continually increasing, adequate protection of exhibits from extraneous DNA is becoming essential. This is an issue that must be addressed if inadvertent contamination of exhibits following the commission of a crime is to be prevented.

Following collection of the sample, the sample should then be packaged or placed in a container that prevents sample loss (eg sealed paper bags, or screw-capped jars for very small specimens that could be lost through other packaging). The exhibit should then be sealed and labelled.

If a sample cannot be picked off an object, it may be possible to remove the sample by cutting it off the surface (eg cutting out a piece of carpet, cutting off a section of wallpaper), packaging it, and labelling and sealing the container.

If neither of the above approaches is appropriate, samples can be removed by swabbing the surface, or by applying adhesive tape. For example, pressing adhesive tape onto the surface and then slowly peeling it off can collect bloodstains on surfaces. Blood will adhere to the tape, and this can be pressed onto a solid surface for support and transport (eg hard plastic sheet/card), packaged, sealed and labelled and subsequently analysed in the laboratory.

Alternatively, blood can be swabbed from the surface using a sterile swab or cotton cloth. The swab/cloth should be dampened with sterile water, and then rubbed over the surface containing the biological material. Once this has been done, the swab/cloth should be air dried to prevent mould growing. This can be difficult to achieve at a crime scene. However, the inclusion of small silica gel packets with the packaged swab/cloth, or careful drying of the swab/cloth as soon as practicable after collection can assist. If this is not possible, the forensic laboratory should be made aware that they may be receiving 'wet' exhibits for analysis, so appropriate arrangements can be made to dry the sample once it has been received.

1.2 Processing of DNA samples

DNA samples that have been collected at crime scenes and from reference samples from individuals can be processed in the forensic laboratory using a myriad of currently available multiplex DNA testing systems. In performing this testing the ultimate objective is to determine whether particular individuals have contributed their DNA to the exhibits, or whether they can be excluded as being contributors to any DNA detected.
DNA testing systems can determine the DNA variants present at one DNA locus (ie singleplex systems) at a time or at more than one locus (multiplex systems) at the same time.

Forensic laboratories prefer to use multiplex systems as they offer greater discrimination to singleplex systems. They also offer time and resource savings as multiple DNA sites can be analysed at one time instead of sequentially.

In December 1997, Australian forensic biology laboratories jointly decided that the Applied Biosystems multiplex Profiler Plus™ was the system of choice for future DNA profiling in Australia (refer to section 2.1). There were other DNA technologies available at that time (eg restriction fragment length analysis), however, these methods required more sample, were less discriminating, and could not test as many DNA loci at one time as PCR based DNA analysis technologies.

Also, at this time the Applied Biosystems multiplex Profiler Plus™ was the most discriminating DNA typing system available. This decision to use this system was prompted for several reasons, amongst which was the impending development of a national DNA database, but in addition the advantages of bulk purchasing arrangements from consumeables suppliers was also relevant.

In the intervening time since December 1997, whilst newer PCR multiplex DNA profiling systems have been developed that offer increased discrimination over Profiler Plus™ (eg Applied Biosystems Identifier, Promega GenePrint® PowerPlex™ 16 System) Australian forensic laboratories have not uniformly adopted these for use. Some laboratories have chosen to validate these systems for use in their hands as an adjunct to Profiler Plus™.

There are a number of reasons why laboratories continue to use Profiler Plus™. One is that movement to another system would necessitate retesting of all previously tested samples if those samples were to be used in a DNA database. This would be considerably resource intensive. In addition, the increased discrimination provided by these other multiplex tests is not necessary for the relatively small size of the Australian population given the discrimination power of Profiler Plus™.

Profiler Plus™ population survey data has shown that the most frequently observed Profiler Plus™ DNA profile may only occur once in approximately every 100 million Caucasians. If it is assumed that the Australian population is approximately 20 million, it can be seen that it would be most unlikely that another individual in the Australian population would have the same Profiler Plus™ DNA profile as any person tested. The greater discriminating power that other multiplex systems may provide would only increase the figure from 100 million to another larger figure. Also, many overseas laboratories are either also using Profiler Plus™ or alternative multiplex systems that are able to analyse most of the loci analysed by Profiler Plus™ enabling comparisons of profiles internationally.

There are several steps integral to the DNA profiling process: DNA extraction, DNA quantitation, amplification of DNA, DNA fragment separation, determination of the
sample DNA profile (ie the DNA genotype) from fragment separation data, quality management and accreditation.

The following sections detail these steps in depth.

### 1.2.1 DNA extraction

There are a variety of methods for the extraction of DNA from forensic specimens. Broadly these are of two main types; organic and non-organic extraction methods.

The choice of which particular extraction method to use for a sample is influenced by a number of factors including the quality of the sample, the surface that the sample is present on, sample processing time and laboratory efficiencies, and laboratory validation studies.

#### 1.2.1.1 Non-organic extraction methods

##### 1.2.1.1.1 Chelex

The most commonly used non-organic DNA extraction method for forensic samples is the chelex extraction method (31-33). This is a relatively rapid method of DNA extraction that is usually used for samples that are not severely affected by other contaminants that may influence the success of subsequent steps in the DNA profiling process.

The basic Chelex procedure consists of boiling the sample in a 5% Chelex solution (34), and then adding a fraction of the supernatant to a DNA amplification reaction after DNA quantitation. If a sample is thought to contain many potential contaminants or chelating metal ions, a 20% chelex solution can be used to attempt to increase the purity of the DNA extracted.

Chelex is a chelating resin that has a high affinity for polyvalent metal ions. The Chelex resin is composed of styrene divinylbenzene copolymers containing paired iminodiacetate ions, which act as chelating groups. It has been postulated that the presence of Chelex during boiling prevents the degradation of DNA by chelating (ie binding) metal ions that may catalyse the breakdown of DNA subjected to high temperatures in low ionic strength solutions (35,36), which are the conditions under which Chelex is used.

Chelex can be used to extract DNA from blood, semen, vaginal fluid, hair, saliva, and some tissue types (eg muscle).

The chelex method does not produce a specified amount of purified DNA per extraction. The amount produced is dependent upon the amount in the sample being analysed. As a result of this, different chelex extractions can produce different amounts of purified DNA. Therefore, DNA prepared using the chelex extraction method can be quantitated if required to more accurately determine how much DNA is present and how much should be added to the subsequent DNA amplification reaction.
1.2.1.1.2 Magnetic beads
An alternative method to the Chelex approach is the use of magnetic bead extraction methods. An example of this is the Promega DNA IQ™ system.

This method uses a specific magnetic resin that purifies (ie captures) DNA without requiring the extensive washing to remove reagents that would normally be added to lyse cells containing DNA. The resin is added to the sample and at the end of the extraction process the resin with DNA attached can be separated from other cellular material in the sample by placing the tube near a magnet. The resin is attracted to the magnet, and this allows the removal of unwanted cellular material from the tube.

The resin has a defined DNA capacity in the presence of excess DNA and will only bind a specific amount of DNA. This property is used to isolate approximately 100ng of DNA from a range of liquid blood, stains or swabs. The resin can also be used to isolate DNA present in hair roots. The DNA attached to the resin can then be eluted from the resin to give a DNA concentration of approximately 1ng/ul. As a result, the analyst can bypass a quantitation step typically necessary with other extraction methods.

There is also the potential for this system to be incorporated into robotics for more rapid sample analysis.

1.2.1.1.3 FTA® paper
Blood and saliva samples can also be collected on a substrate known as FTA® paper. FTA® paper is generally used only for reference samples from individuals. It currently has limited application to the collection of samples from crime scenes.

The FTA® card is a filtration matrix that is impregnated with a patented formula. The formula consists of protein denaturants and chelating agents. The treated matrix also prevents bacterial growth and inactivates organisms such as blood borne pathogens. When a sample is applied to the card the DNA is bound to the matrix. The binding ability of the matrix limits the risk of cross contamination and also the degradation of sample.

When a sample is applied to the FTA® card the cell membranes are lysed, the nucleic acids are released and the DNA is entrapped within the fibres of the cards’ matrix. Purification/extraction of the DNA is simple (only requiring a series of wash steps) and amplification is then performed with the FTA® disc remaining within the tube.

This means that DNA quantitation need not and cannot be performed for samples extracted from FTA® paper.

1.2.1.2 Organic extractions of DNA
This method employs a detergent (sodium dodecyl sulphate) to rupture cell membranes and a protease to denature the proteins and expose the DNA, followed by
solvent extraction (phenol chloroform) to remove protein and lipid, and finally removal of the solvent (30,37,38). It is used when samples are degraded or when there may be large amounts of contaminating substances (eg dyes) present that have the potential to adversely affect subsequent steps in the DNA profiling process.

The methodology for some of the tissue/fluid types typically encountered in forensic samples can be summarised as follows:

**Liquid blood:**
White blood cells are purified and lysed using a detergent and the proteins present are denatured with detergent and protease. The protein and other denatured cellular material are removed using organic solvents and the solvents are removed by filtration.

**Bloodstains:**
The dried bloodstains are washed to remove haem (a substance present in blood) and the cellular material left adhering to the substrate (eg fabric) is then denatured using detergent and protease. The digest is then treated as with whole blood.

**Trace DNA samples:**
Trace DNA can be treated as bloodstains without the need to perform the haem removing wash.

**Saliva/buccal cells and semen-free vaginal swabs:**
The cellular material in stains et cetera is denatured using detergent and protease and the digests treated as for whole blood.

**Hairs:**
Hair roots are cut off and treated with detergent, protease and reducing agents. The digests are then treated as for whole blood.

**Seminal stains or semen/vaginal fluid mixtures:**
Swab/stains are first treated with detergent and protease. The digest is then centrifuged and the supernatant removed and treated as for whole blood to isolate the ‘female’ fraction. The pellet is then treated with detergent, protease and reducing agent that breaks down the proteins bound to the DNA in spermatozoa. The second digest is separately treated as for whole blood to isolate the ‘male’ fraction. Note that the ‘male’ fraction contains cellular material principally from spermatozoa, and the ‘female’ fraction contains material from non-spermatozoa cell types.

**Blood/semen mixtures:**
Swabs/stains are washed to remove the cellular material left adhering to the fabric. The solution is treated as with bloodstains to isolate the ‘female’ or blood fraction from the spermatozoa, which are recovered by centrifugation. The pellet is then treated with detergent, protease and reducing agent and this second digest is separately treated as with whole blood to isolate the ‘male’ fraction.

**Bone:**
The bone is cleaned of surface contaminants. A hacksaw is used to produce shavings of the interior of the bone. The bone tissue is digested with protease and detergent. A
high concentration of EDTA is used to chelate excess calcium ions. DNA is extracted from the digested material using phenol and chloroform as for blood and other tissues. The method is based on that of Hagelburg (30).

1.2.1.3 Further purification of extracted DNA
DNA that has been extracted by the Chelex method can be subjected to organic extraction methods to further purify it. This may be required when the original sample has been on a surface heavily contaminated with dyes or other contaminants that were not fully removed by the Chelex method.

In addition, there are a variety of other methods available to ‘clean up’ DNA that has been extracted from samples. For example, DNA can be passed through Centricon tubes (Millipore), which essentially act as filters. The DNA is placed into the tube and other liquid added to the tube. On centrifugation of the tube, contaminants pass through a micro sieve, and DNA is retained behind. The retained DNA is therefore more pure than it was prior to this treatment.

Another type of tube that can be used to further purify DNA in a similar manner is the Qiagen QIAquick PCR Purification protocol (Qiagen). The QIAquick kit and system combines spin-column technology with the selective binding properties of a silica-gel membrane. Buffers provided in the kit are optimized for efficient recovery of DNA and removal of contaminants. DNA adheres to the silica membrane in the presence of high salt while contaminants pass through the column. Impurities are washed away and the purified DNA is eluted with buffer or water.

1.2.2 DNA Quantitation
Once DNA has been extracted, it may be necessary to estimate how much DNA is present in an extracted sample. The efficiency of a Polymerase Chain Reaction (PCR) amplification is influenced by the quality (degree of degradation), purity, and total quantity of DNA in a sample. Lack of amplification is usually due to highly degraded DNA, the presence of PCR inhibitors, insufficient DNA quantity, or any combination of these factors.

1.2.2.1 Quantiblot
There are a number of ways in which DNA can be quantitated in preparation for use in a PCR amplification reaction. One commonly used method is the quantification of human DNA using the ‘Quantiblot Human DNA Quantitation Kit’ (Roche) (39). If this method determines that sufficient DNA is present in the extracted sample, then lack of amplification is most likely due to PCR inhibitors or severe degradation of DNA.

Quantitation of samples determines if there is sufficient DNA for amplification. In addition, adding the smallest volume of DNA extract necessary for successful amplification can minimize PCR inhibition. Lastly, by using the minimal volume of extracted DNA for PCR, the number of genetic marker tests or repeat analyses that can be performed is maximised.
DNA quantitation is particularly important for Profiler Plus™ amplifications, where optimal results are obtained using a range of typically 1-2.5ng of starting DNA. Adding more DNA can result in too much PCR product, such that the dynamic ranges of the instruments used to detect and analyse PCR product is exceeded. Adding too little starting DNA means that insufficient amplified DNA product may be synthesised to be detectable by the instruments.

The Quantiblot process is specific to human and higher primate DNA. It consists of the hybridisation of a biotinylated oligonucleotide probe to DNA, which is immobilised on a nylon membrane. The probe is specific to a tandem repeat at the DNA locus D17Z1. Short tandem repeat DNA sequences are explained in more detail in section 1.2.3 Amplification of DNA.

The subsequent binding of horseradish peroxidase-streptavidin enzyme conjugate (HRP-SA) to the bound probe allows for either colourimetric or chemiluminescent detection.

In the colourimetric detection, the oxidation of 3,3’,5,5’-tetramethylbenzidene (chromogen TMB) catalysed by the HRP-SA results in the formation of a blue precipitate on the membrane. The amount of blue precipitate is proportional to the amount of DNA.

For chemiluminescent detection, the oxidation of a luminol-based reagent catalysed by HRP-SA results in the emission of photons that are detected on standard autoradiography (X ray) film. This process is called enhanced chemiluminescence.

![A Diagrammatic Representation of the Quantiblot Process](image)

**Figure 1: A Diagrammatic representation of the Quantiblot process**

Results are interpreted by comparing the signal intensity of the DNA test sample to the signal intensity obtained for DNA Standards that are also included in the kit.

Two DNA Calibrators that are included in the kit are used to provide DNA of a known concentration to verify that the DNA Standards were correctly applied and are providing correct results for the test samples.
On occasion, particularly with trace DNA samples, dye from the sample can be present in the extracted DNA that in turn is bound to the Quantiblot membrane. This colouration may mask colouration that is present due to DNA. This can make it difficult to accurately estimate the amount of DNA in a sample in the colourimetric version of this test. For this reason this method generally only provides an indicator of the amount of DNA present rather than a precise measurement.

1.2.3 Amplification of DNA

DNA is a complex structure made of sequences of molecules called bases. There are four types of base: adenine (A), guanine (G), cytosine (C) and thymine (T). A person’s DNA (ie their genome) is unique because the sequence of bases comprising their DNA molecules is unique. Certain parts of the sequence pertain to genes that in turn determine a person’s physical characteristics. These areas therefore have a known function.

DNA preferentially exists in a double stranded conformation, known as a double helix. Since adenine preferentially binds to thymine, and guanine preferentially binds to cytosine, a DNA strand will be bound to another strand which contains the complementary sequence of bases to that found in the first strand. For example, if a length of DNA has the sequence AATCGTTACCG, the complementary sequence will be TTAGCAATGGC.

Throughout the human genome there are certain DNA sequences known as short tandem repeats (STRs). These are sequences of DNA that are repeated numerous times. The reason why these are present in DNA is not known; however, whilst they appear to have no known current function, they are of particular interest for forensic testing. This is because different individuals have different numbers of the tandemly repeated core base sequences allowing the discrimination of individuals by DNA testing.

The details of the repeated sequences amplified by the Profiler Plus™ kit are detailed in table 1.

The section of the genome that is amplified is determined by the specificity of pairs of oligonucleotide primers. The Applied Biosystems AmpF/STR Profiler Plus™ PCR Amplification kit includes reagents to amplify nine short tandem repeat (STR) loci: D3S1358, vWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317 and D7S820 and also the sex indicating marker Amelogenin. The details of each are shown in the table 1:
<table>
<thead>
<tr>
<th>Locus</th>
<th>Chromosome location</th>
<th>Repeat Sequence$^b$</th>
<th>Size Range (bp)$^a$</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>D3S1358</td>
<td>3p</td>
<td>TCTA (TCTG)$_{1,3}$</td>
<td>114-142</td>
<td>Li et al. 1993 (40)</td>
</tr>
<tr>
<td>VWA</td>
<td>12p12-pter</td>
<td>TCTA (TCTG)$_{3,4}$</td>
<td>157-197</td>
<td>Kimpton et al. 1992 (41)</td>
</tr>
<tr>
<td>FGA</td>
<td>4q28</td>
<td>(TCTA)$_n$</td>
<td>219-267</td>
<td>Mills et al. 1992 (42)</td>
</tr>
<tr>
<td>D8S1179</td>
<td>8</td>
<td>(TCTR)$_n$</td>
<td>128-168</td>
<td>Oldroyd et al. 1995 (43)</td>
</tr>
<tr>
<td>D21S11</td>
<td>21</td>
<td>(TCTG)$_n$</td>
<td>189-243</td>
<td>Sharma and Litt 1992 (44)</td>
</tr>
<tr>
<td>D18S51</td>
<td>18q21.3</td>
<td>(AGAA)$_n$</td>
<td>273-341</td>
<td>Urquhart et al. 1995 (45)</td>
</tr>
<tr>
<td>D5S818</td>
<td>5q21-31</td>
<td>(AGAT)$_n$</td>
<td>135-171</td>
<td>Hudson et al. 1995 (46)</td>
</tr>
<tr>
<td>D13S317</td>
<td>13q22-31</td>
<td>(GATA)$_n$</td>
<td>206-234</td>
<td>Hudson et al. 1995 (46)</td>
</tr>
<tr>
<td>D7S820</td>
<td>7q</td>
<td>(GATA)$_n$</td>
<td>258-294</td>
<td>Green et al. 1991 (47)</td>
</tr>
<tr>
<td>Amelogenin</td>
<td>X:p22.1-22.3</td>
<td>N/A</td>
<td>107</td>
<td>Sullivan et al. 1993 (48)</td>
</tr>
<tr>
<td></td>
<td>Y:p11.2</td>
<td></td>
<td>113</td>
<td></td>
</tr>
</tbody>
</table>

$a$ The size range is the actual base pair size of sequenced alleles contained in the AmpF/STR Allelic Ladders. For example a bp size of 200 means that the DNA fragment has 200 bases in it, and is paired in a complementary fashion to another strand also of 200 bases in length, giving 200 base pairs (bp).

$^b$ R represents either an A or a G

$n$ indicates that this particular part of the repeat DNA sequence can be repeated any number of times, where this number is known eg (ATG)$_n$ where $n=1$ to 5, means that the sequence could be ATG, ATGATG, ATGATGATG, ATGATGATGATG or ATGATGATGATGATG.

Dyes: 5-FAM for D3S1358, VWA, FGA
JOE for Amelogenin, D8S1179, D21S11, D18S51
NED for D5S818, D13S317, D7S820

Table 1: AmpFlSTR Profiler Plus™ PCR Amplification Kit STRs analysed

Depending upon the concentration of DNA in the sample to be amplified, a volume of sample that contains approximately 1-2.5ng DNA should be added to a tube. This volume should be 20ul maximum.

Thirty microlitres of a PCR ‘master mix’ should then be added to each tube. The master mix is made by adding 20:1:11 volumes of AmpF/STR PCR Reaction mix: AmpliTaq Gold DNA Polymerase: AmpF/STR Profiler Plus™ Primer Set. The
chemicals that make up the master mix are obtained from the AmpF/STR Profiler Plus™ PCR Amplification Kit.

In addition, a tube containing all reagents except DNA is prepared. This tube is known as an amplification negative sample. Its purpose is to demonstrate that DNA is absent from all reagents used in the amplification reaction, and therefore that amplified DNA detected in samples is in fact from the actual sample and not from DNA that may have been present in reagents.

Another sample is also amplified at the same time. This sample is an amplification positive sample, and is found in the AmpF/STR Profiler Plus™ PCR Amplification Kit. This sample is of a known genotype and is used to confirm that for a particular amplification batch that the correct genotypes are obtained. This can be inferred if the amplification positive sample produces the known expected genotype on analysis.

Placing the amplification reaction tubes in a thermalcycler commences the amplification process. This is an instrument that takes tubes through specified cycles of alternating temperatures. The following table illustrates the temperature cycling parameters that are used for Profiler Plus™ amplifications:

<table>
<thead>
<tr>
<th>Step</th>
<th>No. of Cycles</th>
<th>Temp°C</th>
<th>Time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Incubation</td>
<td>1 Hold</td>
<td>95</td>
<td>11</td>
</tr>
<tr>
<td>Amplification</td>
<td>28 cycles of: denaturation annealing extension</td>
<td>94</td>
<td>59</td>
</tr>
<tr>
<td>Final Extension</td>
<td>1 Hold</td>
<td>60</td>
<td>45</td>
</tr>
<tr>
<td>Hold File</td>
<td>1 Hold</td>
<td>10</td>
<td>forever</td>
</tr>
</tbody>
</table>

Table 2: Amplification parameters for AmpF/STR Profiler Plus™ amplification reactions

The mechanism by which amplification occurs is detailed in figure 2. Each amplification cycle doubles the amount of DNA produced. For Profiler Plus™, this amplification is occurring at all 10 loci at the same time and in the same reaction tube.

The 10 sets of primers (one primer pair per locus amplified) are short lengths of DNA complementary to a region near the target sequence (locus). The primers have a fluorescent molecule incorporated into their structure, and it is the fluorescent molecule that is used in the fragment separation steps of DNA profiling to identify the particular DNA sequences or variants at each locus that have been amplified.
1.2.4 Fragment Separation

Once the amplification process has been completed the fragments must be separated so the particular DNA variants (ie alleles) present can be determined. This is done typically either by separating the fluorescently labelled fragments via electrophoresis or by injection of the sample through a capillary. This is known as multicomponent analysis.

Multicomponent analysis is a process that separates the four different fluorescent dye colours used in Profiler Plus™ analysis into distinct spectral components. The three different dyes used in the AmpFlSTR Profiler Plus™ PCR Amplification Kit are the NHS-ester dyes known as 5-FAM, JOE, and NED. A fourth dye, ROX, is used with an internal lane size standard. Note that there are other dyes that can be used for other multiplex systems.

Each of these dyes is incorporated into the amplified DNA products generated in the PCR process. Each of the dyes emits its maximum fluorescence at a different wavelength. This light emission is initiated by shining a laser onto the fluorescent fragments. During data collection which can be performed on a variety of fragment separation instruments (eg the ABI Prism®310, 377, 3100, and 3700 instruments), the fluorescent signals generated by laser excitation are separated by a diffraction grating according to their wavelength, and projected onto a Charged Couple Device (CCD) camera in a predictably spaced pattern.

For Profiler Plus™ analyses, 5-FAM emits at the shortest wavelength and is detected as blue light, followed by JOE (green), NED (yellow) and ROX (red), listed in order of increasing wavelength.
Although each of these dyes emits its maximum fluorescence at a different wavelength, there is some overlap in the emission spectra between the four dyes. The goal of multicomponent analysis is to isolate the signal from each dye so that, for example, signals from 5-FAM labelled products are displayed in the electropherogram for blue, but not in those for green, yellow or red. The electropherogram is a pictorial representation of the data collection.

The precise spectral overlap between the four dyes is measured by analysing DNA fragments labelled with each of the four dyes in separate lanes of a gel or in separate injections on a capillary. These dye-labelled fragments are called matrix standards.

ABI Genescan® Analysis Software then analyses data from each of these four samples and creates a matrix file. The matrix file contains a table of numbers with four columns and rows (refer to figure 3)

<table>
<thead>
<tr>
<th></th>
<th>Reactions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
</tr>
<tr>
<td>B</td>
<td>1.0000</td>
</tr>
<tr>
<td>G</td>
<td>0.5916</td>
</tr>
<tr>
<td>Y</td>
<td>0.3137</td>
</tr>
<tr>
<td>R</td>
<td>0.1211</td>
</tr>
</tbody>
</table>

Table 3: Matrix file table example

These numbers are normalized fluorescence intensities and represent a mathematical description of the spectral overlap that is observed between the four dyes.

The rows in the table represent virtual filters and the columns represent the dye-labelled PCR products, indicated as ‘Reactions’. The top lefthand value, 1.0000, represents the normalised fluorescence of the blue (5-FAM labelled) PCR products in the blue filter. It follows that all matrix tables should have values of 1.0 on the diagonal from top left to bottom right, as shown in figure 3.

The other values in the table should all be less than 1.0. These values represent the amount of spectral overlap observed for each dye in each virtual filter. For example, the values in the first column reflect quantitatively the amount of blue dye detected in each virtual filter.

The matrix values obtained will vary between different instruments and at different times for the same instrument, and can vary between different run conditions on a single instrument. A matrix file must be made for each instrument and for a particular set of run conditions.

The appropriate matrix file can be applied to data on subsequent capillary runs or gels on the same instrument, as long as the gel formulation and running conditions are constant from run to run. This is because spectral overlap between the four dyes is reproducible under constant run conditions.
Multicomponent analysis is achieved automatically by the ABI Genescan® Analysis Software, which applies a mathematical matrix calculation (using the values in the matrix file) to all sample data (49).

At the completion of data analysis through the ABI Genescan® Analysis Software, the raw data collected from the fragment separation instruments is converted into tabular information that details when particular DNA fragments were detected by the instrument. It also provides information about the relative size and amounts of each of the fragments detected.

Following this, the data is then imported into another software package known as ABI Prism™ Genotyper® software. This software application enables the analysis and interpretation of DNA fragment size and quantitation data by converting it into user defined information. Typically, this results in data that provides allele designations, and also provides information about the quantity of each particular DNA fragment present.

A typical result for a sample that contains DNA from one source is presented in figure 3.

Figure 3: Electropherogram showing the Genotyper plot of a sample containing DNA from a single source

Figure 3 illustrates some key characteristics of samples that contain DNA from a single source. Each of the sets of peaks show the DNA alleles (variants) present at each of the 10 loci analysed using the Profiler Plus™ multiplex. It can be seen that the alleles are present either as pairs, or as a single peak at a locus. A person will have either two different DNA alleles at a locus (example: D5S818) or one DNA allele at a locus (example: D7S1358). Since an individual receives half of their DNA from each parent, where two different alleles are present, this means that each of that person’s
parents contributed a different allele to the child. When one allele is present, this means that each of the person’s parents contributed the same allele to the child.

Samples containing mixtures of DNA from more than one individual are different. An example of an electropherogram showing the Genotyper plots of a sample containing DNA from a mixture can be seen in figure 4.

Figure 4: Electropherogram showing the Genotyper plot for a sample containing a mixture of DNA from more than one source.

Figure 4 shows that where DNA from more than one individual is present in a sample, that it is possible to observe more than two alleles at a locus, for example locus D5S818. At this locus there are 4 alleles present. This means that there is DNA from at least 2 individuals in this sample.

Interpretation of mixture profiles is more difficult than interpretation of single source profiles. This is typically because the amount of DNA present from the different individuals contributing to the one sample may vary. For example, a sample may contain DNA from two individuals, and there may be similar amounts of DNA from each person present. In this case, it would be reasonable to expect that the peak heights obtained for each allele in the mixture at a particular locus would be approximately the same. However, for a sample that contains DNA from two individuals where the amount of DNA present from one individual is considerably different to that from the second person, extreme differences in allele peak heights for the alleles from each person may be observed. In some cases, the minor component of a mixture may be so minor that poor DNA amplification of this component may occur and it may become very difficult to detect its presence.
1.2.5 Quality Management and Accreditation

World wide, forensic laboratories are increasingly subscribing to accreditation programs.

Reeder (50) reviewed the history of DNA-based human identification. Events that occurred particularly in the USA in the early 1990’s contributed towards recognition of the need for standards for DNA profiling.

In Australia, the most widely used program is that offered by the National Association of Testing Authorities, Australia (NATA). This is the national organisation for conformity assessment of technical operations such as laboratories, inspection bodies and reference materials producers. By way of a Memorandum of Understanding, the Commonwealth Government recognises NATA as the sole national accreditation body for establishing competent laboratory practice. NATA also represents Australia in the International Laboratory Accreditation Cooperation (ILAC), the Asia Pacific Laboratory Accreditation Cooperation (APLAC) and on the Organisation for Economic Cooperation and Development Panel on Good Laboratory Practice.

NATA aims to provide, in the national interest, an accreditation service which meets the needs of stakeholders, and also facilitates the recognition and acceptance of their products and services. NATA also promotes the science and practice of accreditation to enhance the acceptance of Australian products and services both in Australia and overseas.

The cornerstone of NATA accreditation is peer assessment whereby peers of the laboratory or facility under review ‘assess’ the technical competence of that facility.

Laboratories are required to meet criteria detailed under AS ISO/IEC 17025 ‘General Requirements for the Competence of Testing and Calibration Laboratories’ (51).

In addition, forensic laboratories are required to also meet ISO/IEC 17025 ‘Application Document Supplementary Requirements for Accreditation in the Field of Forensic Science’ (52). This latter document in particular contains many specific requirements for forensic laboratories wishing to undergo accreditation. Amongst the many requirements are the following: the organisation itself, quality systems, document control, subcontracting of tests and calibrations, service to the client, corrective actions, preventive actions, control of records, internal audits, management reviews, personnel, accommodation and environmental conditions, test and calibration methods and method validation, measurement traceability, sampling, handling of test and calibration items, assuring the quality of test and calibration results, and reporting results.

Annex 1 of ISO/IEC 17025 ‘Application Document Supplementary Requirements for Accreditation in the Field of Forensic Science’ details the additional criteria that must be met by DNA laboratories seeking accreditation.

The culmination of these accreditation requirements is a quality management system that requires and supports quality work output and the immediate and efficient rectification of errors should they occur.
1.3 **Storage of DNA samples**

In terms of physical storage, DNA is a very robust molecule and can survive relatively undegraded, in extreme environmental conditions, and sometimes for many years. However, degradation is minimised by storing samples under more suitable conditions.

Extracted DNA and amplified product DNA will undergo the slowest rate of degradation stored at –20 to –80°C.

DNA present in crime samples and exhibits can be stored for extended periods of time at room temperature without significant DNA degradation, provided the sample is stored dry and in a non-humid environment. Reduced degradation will occur if these samples are stored at lower temperatures than room temperature. However, the logistics of storing crime scene exhibits given their variable size and nature means that most exhibits are stored at room temperature.

Reference DNA samples from individuals can be successfully stored for long periods at room temperature for some sample types (eg blood or buccal samples on FTA paper or isocode stix), or at lower temperatures for other sample types (eg liquid blood or other liquid body fluids, muscle).

However, in terms of the philosophical aspect of storage of samples for future use, this is frequently controlled and limited by legislation. For example, the Victorian Crimes Act 1958 (53), and the Crimes (DNA Database) Act 2002 s464ZGJ and s464ZGK (54) list the various circumstances under which samples taken under these acts can be used. Essentially they can only be used for inclusion on a DNA database or in the investigation of a criminal offence. There are no provisions for the use of samples taken under this legislation for any other purpose such as medical research for example.

The legislation also allows for samples to be made available for independent testing when there is sufficient material to be analysed both in the investigation of the offence and on behalf of a person from whom a sample has been taken in relation to that offence (refer to s464ZC Analysis of material found at scene of offence etc. (53)). Therefore, samples may be required to be stored for long periods (years) and potentially indefinitely.
The establishment and use of a DNA database

DNA databases provide an ability to potentially identify possible offenders to unsolved crime and to link crimes that were previously unknown to be linked. In this sense, DNA databases enable pro-active policing and have the potential to increase the solving of crime.

2.1 The establishment of DNA databases

2.1.1 General

In 1986, Dawn Ashworth was murdered in Leicestershire, UK. The police involved in investigating this murder and another seemingly linked murder in 1983, decided to use the new technique of DNA profiling in their investigation. In conjunction with the then UK Central Research Establishment (now the Forensic Science Service), the police created an informal DNA database by collecting blood samples from all the males in the town where Dawn Ashworth lived. This was the first use of a ‘DNA database’ to solve a crime.

In subsequent years many other countries have evaluated the potential use or creation of DNA databases.

2.1.2 Australia

In 1990, the (then) Australian Police Minister’s Council (APMC) received a report from Dr P Easteal entitled ‘The Forensic Use of DNA Profiling in Australia: Need for a National DNA Database’. In 1993, the APMC received a report from Mr A Ross (Director, National Institute of Forensic Science) entitled ‘Considerations of the Easteal Report’. The APMC resolved to note the Ross Report and endorse its recommendations.

The APMC subsequently commissioned a working party under the Chairmanship of Mr Justice JH Phillips, Chief Justice of Victoria and Chairman of the National Institute of Forensic Science (NIFS) to further consider the recommendations of the Ross Report. The report recommended, amongst other recommendations, that a small working party be set up to determine whether Australian State/Territory legislation to obtain body samples should be standardised, and to make recommendations on legislation relating to privacy issues and database integrity.

An interim report and recommendations of the Working Party were considered at the 29th Meeting of the APMC in November 1995. The report had two main purposes. The first was to provide information to the NIFS Board of Control, the Senior Officers Group (SOG) and the APMC on the progress made by the Working Party in defining issues relative to the establishment of a national DNA database. The second was to provide a chance for other groups such as participating laboratories and the Privacy and Human Rights Commissions to make comment on the interim recommendations.

Following this, in August 1997, a meeting of senior police and forensic scientists was convened to discuss issues relating to a national convicted offender DNA database. The keynote speaker was Chief Constable Ben Gunn, Cambridgeshire Constabulary, UK. Mr Gunn represented the Association of Chief Police Officers (ACPO) in the
development and maintenance of a national convicted offender DNA database in the UK.

A key outcome from this meeting was to seek to have the issue of the database placed in the agenda for the Crime Commissioner’s Conference in October 1997. Another key outcome was that a nationally common system for DNA profiling should be adopted (A Ross, personal communication).

Subsequently, in November 1997, senior managers of forensic laboratories selected Profiler Plus™ as the nationally common system for DNA profiling

2.1.2.1 The establishment of the Australian National Criminal Investigation DNA Database (NCIDD)

Background:
In 1998 the Commonwealth Government set aside $50 million to improve the solution of crime through the useful aggregation of jurisdictional policing data into a common database(s).

The CrimTrac Agency was established on 1 July 2000. The Agency was established by the Commonwealth Government to oversee the expenditure of monies and develop a common view of policing data. CrimTrac will allow Australian police services to take advantage of the dramatic opportunities opened up by recent advances in forensic science, information technology and communications.

The CrimTrac Agency was established as an Executive Agency under the Commonwealth Public Service Act in the Attorney-General's portfolio on 1 July 2000. The Chief Executive Officer of CrimTrac reports to the Federal Minister for Justice and Customs, and the Agency must comply with all Commonwealth legislative, financial and administrative arrangements.

The Agency is underpinned by an Inter-Governmental Agreement signed by all Australian police ministers. The Australasian Police Ministers' Council is responsible for defining the Agency's strategic directions and key policies, setting new initiatives, and appointing members to the CrimTrac Board of Management. The Board is generally responsible for the overall management of the Agency. The Chief Executive Officer also reports to the Board on the programs and performance of CrimTrac.

One of the key initiatives was the establishment of a National Criminal Investigation DNA Database (NCIDD) to supplement existing jurisdictional databases. The Prime Minister formally launched Phase 1 of NCIDD, a central database with matching engine, on the 20th June 2001.

Brief History:
Development of NCIDD commenced in November 1998. CrimTrac consulted with the jurisdictions to establish the scope of NCIDD and the CrimTrac Board of Management (BOM) established a DNA Steering Committee comprising senior police and forensic laboratory managers. The structure of NCIDD was developed under the guidance of the steering committee.
The model developed for NCIDD is a distributed model comprising a central, national database supported by a jurisdictional interface for both quality assurance and to enable selection of the most appropriate DNA profiles to be sent to the central database.

**Fundamental requirements of NCIDD:**
A number of principles were agreed for NCIDD:

1. That the operation of NCIDD would rest with the forensic laboratories – The major concern was that identified matches needed to be forensically confirmed for accuracy before advising the police;

2. That individuals could not be identified from the information held on the central database;

3. That NCIDD would be capable of interface with existing police information systems (eg COPS, PROMIS);

4. That the system would be secure including logged audit trails;

5. That the system would be able to accommodate variations in jurisdictional matching rules as set out in different legislation;

6. That destruction dates could be accommodated;

7. That the system could accommodate mixtures of DNA profiles; and

8. The system could accommodate future changes in DNA profiles (specifically more loci), including profiles from external sources (eg CODIS), and be scalable.

**The NCIDD system:**
To achieve all the above criteria the system developed for NCIDD is:

1. A purpose built central database containing a matching engine, based on open computer systems, which accommodates:
   a. The legislative matching criteria;
   b. Destruction date management;
   c. Audit trail logging; and
   d. Mixtures – (awaiting advice of forensic community).

2. A jurisdictional laboratory interface for:
   a. Automating the upload of profiles to the central database;
   b. Accommodating future possible linking to existing police systems;
c. Audit trail logging; and

d. Assuring the quality of DNA profiles before upload to the central database.

3. A secure, closed communications network linking each jurisdictional laboratory with the central database (implemented and managed by the Australian Federal Police).

Phase 1 – the central database is being readied for the release of Version 2 of the software, which incorporates a number of enhancements to assist the user manage data related to matches.

Phase 2 – the product of a commercial supplier is currently under trial to assess the product’s acceptability for the jurisdictional interface.

The secure communications network is in place.

The system has been designed so that the two phases are independent. This has enabled the development of the central database while still assessing the requirements of the jurisdictional interface.

Figure 5 illustrates the current system.

**NCIDD Network**

- Central database in Canberra
- Jurisdictional forensic labs linked to database via secure AFP network
- Majority of labs connected through ISDN lines
- All data on network is encrypted
NCIDD

Profiles uploaded in bulk or individually through PC and results viewed via web browser

Figure 5: NCIDD network

The legislation Issue:
The major issue affecting the optimum performance of NCIDD is variability of DNA legislation between jurisdictions. Nine separate pieces of legislation govern the use of DNA profiles in criminal investigations, within Australia. An assessment of existing DNA legislation in early 2001 identified variation between current legislations would result in cross-jurisdictional matching being well short of optimal, with less than 50% of possible matching (based on categories defined in legislation such as crime scene, offender, suspect, volunteers, etc) being able to occur between jurisdictions.

Since that assessment, considerable activity has been undertaken by a number of jurisdictions to amend legislation to increase cross-jurisdictional matching. The Commonwealth, NSW, ACT and Tasmania have implemented legislation based on the Model Criminal Code. Recently WA has implemented legislation that aligns with the model code. Victoria has recently made amendments to its legislation to align with the model code while SA has introduced into its Parliament, model code amendments to its current legislation. Only Queensland and NT have yet to consider similar amendments to their legislation.

Encouragingly, at the April 2002 Australian Leaders Summit on Transnational and Multi-jurisdictional crime, leaders agreed (inter alia):

- *To modernise the criminal law by legislating in the priority areas of model forensic procedures (during 2002)....

and
• To enhance capacity in each jurisdiction for the collection and processing of samples to create DNA profiles, and the uploading of profiles onto the national DNA database.

For cross-jurisdictional matching to occur on NCIDD, each jurisdiction must undertake a number of activities:

• Arrangements must be put in place between two jurisdictions to share information
• Regulations must be issued for each piece of legislation recognising other jurisdictions’ legislation; and finally
• Agreements need to be established between the jurisdictions indicating which categories (or indexes) can be legally matched (due to definitional differences in the categories).

Considerable work has been done by all jurisdictions in relation to the first two dot-points and work is commencing on the third dot-point.

How NCIDD operates:
Profiles forwarded to the central database have only the profile and some sample/case management information such as unique case/sample ID, destruction date, and category (typically about 500 characters of data). No unique identifying information is contained (such as name, address, etc), only the jurisdictions hold this information.

Profiles uploaded from the laboratory to the central database are automatically matched against the profiles already on the database, in accordance with prescribed matching rules (based on legislative requirements).

The forensic scientists in each jurisdiction then view potential ‘linked’ profiles via a web browser, operating across the closed, secure AFP network to the NCIDD. Contact is then made with a forensic scientist in the other jurisdiction and the process of validating the match using laboratory records proceeds. Once each laboratory has confirmed a valid match, the forensic scientists will advise the police with full details of the match.

The system can also be used for intra-jurisdictional matching.

The system also manages destruction dates of profiles by removing all traces of nominated profiles, and associated matches, when the destruction date falls due. A facility is included in the system to enable jurisdictions to review destruction dates (eg due to an on-going court case, or where a suspect is subsequently convicted).

Governance of NCIDD:
Overseeing the operation of NCIDD is a CrimTrac User Advisory Group (UAG) reporting to the CrimTrac Board of Management. The UAG will monitor the performance of NCIDD and recommend enhancements to the system.

The UAG comprises:
• senior police representatives
senior forensic laboratory managers

a representative from the National Institute of Forensic Science

a user representative representing the Biology Special Advisory Group of SMANZFL (Senior Managers of Australian and New Zealand Forensic Laboratories)

a CrimTrac representative

Figure 6 illustrates the governance model for NCIDD.

A common MOU has also been developed between CrimTrac and a jurisdiction, setting out the responsibilities of both parties in relation to the use and operation of NCIDD

Some key dates relating to DNA databases:
1989 Australia's first court case involving DNA evidence. In an ACT court, Desmond Applebee is convicted of three counts of sexual assault. Applebee changes his defence from "I wasn't there" to "the woman consented" after a blood sample matches him to DNA extracted from blood and semen on the victim's clothes.
1989 In Victoria, police secure the conviction of George Kaufman who raped sixteen women over a four year period in Melbourne's south eastern suburbs. Confronted with DNA evidence, Kaufman confesses.

1989 The Federal Government and several States and Territories begin developing regulatory standards for DNA collection and handling procedures.

1992 National Institute of Forensic Science commences operations. Amongst its roles are the development of national standards of quality control and accreditation of forensic laboratories throughout Australia.

1995 The world's first national DNA database commences operations in the UK on 10 April 1995.

1996 In the USA, mitochondrial DNA evidence is used in a court for the first time. Paul Ware is convicted of the rape and murder of a four year old girl after mitochondrial DNA profiling matches him to a hair found on the body of the child.

1996 Rodney Winters is convicted of the rape and murder of a woman at South Australia's Edinburgh Air Force base 14 years earlier. After DNA profiling matches him to semen found on the dead woman, Winters confesses.

1997 Police services endorse the establishment of a national criminal DNA database and form a working party.

1997 Victoria becomes the first jurisdiction in Australia to enact legislation regulating the use of a DNA database.

1998 In the USA, the FBI sets up the National DNA Index System, enabling city, county, state and federal law enforcement agencies to compare DNA profiles electronically.

1998 Australian forensic laboratories agree to a common national standard for obtaining DNA profiles.

1998 The Australian Federal Government commits $50m to establish CrimTrac, with a national DNA database as a central element.

1999 Victorian police obtain first 'cold' hit from state DNA database - the DNA profile of convicted thief Wallid Haggag is matched to blood found in a car used in a burglary for which he was not previously a suspect.

2000 Following the sexual assault of a 91 year old woman in Wee Waa, New South Wales police take DNA samples from 500 local men aged between 18 and 45. Stephen Boney confesses shortly afterwards, before his sample can be analysed.

2000 In the UK, the Forensic Science Service announces that the number of DNA profiles of suspects and convicted criminals on the national DNA database has reached one million or roughly one third of the estimated criminally active population.
2001 First reported Australian case of DNA evidence assisting with the exoneration of a convicted offender. Mr Frank Button is released from a Queensland jail where he had served 256 days of a three year sentence for the rape of a 13 year old girl. Mr Button also spent 101 days on remand prior to sentencing.

Overseas DNA Databases – some information:

United Kingdom

The world's first national criminal DNA database was established in the United Kingdom in April 1995, following a Royal Commission into the criminal justice system. It is called the National DNA DataBase (NDNADB)

The NDNADB now holds DNA profiles from more than 1.579 million suspects and from more than 132,000 crime scenes.

In the last seven years, the NDNAB has achieved the following numbers of “hits”:
- Person to Scene – 164,200
- Scene to Scene – 14,628

The database is currently achieving about 1400 matches every week between crime scene samples and suspects. Police have used the database as an investigative tool to help solve a wide range of crimes, as seen in table 4*:

<table>
<thead>
<tr>
<th>Type of Crime</th>
<th>No. of Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Murder / Manslaughter</td>
<td>1047</td>
</tr>
<tr>
<td>Attempted Murder</td>
<td>397</td>
</tr>
<tr>
<td>Supply / Attempt to supply drugs</td>
<td>1433</td>
</tr>
<tr>
<td>Rape</td>
<td>1849</td>
</tr>
<tr>
<td>Indecent Assault</td>
<td>582</td>
</tr>
<tr>
<td>Serious Robbery</td>
<td>1904</td>
</tr>
<tr>
<td>Wounding / Grievous bodily harm</td>
<td>957</td>
</tr>
<tr>
<td>Kidnapping / abduction</td>
<td>307</td>
</tr>
<tr>
<td>Arson</td>
<td>527</td>
</tr>
<tr>
<td>Other serious offences</td>
<td>896</td>
</tr>
</tbody>
</table>

*Source: Chief Constable Ben Gunn, Cambridgeshire Constabulary, UK, 2002

Table 4: UK NDNADB figures

Using the NDNADB, UK police have almost doubled their clearance rate for volume crimes such as house burglary.

New Zealand

New Zealand's national DNA databank commenced operations in August 1996.

The database now holds over 16,000 individual DNA profiles, with approximately 400 more profiles being added each month.
The database also holds over 1,900 crime scene profiles, with approximately 120 more profiles being added each month.

Approximately 36% of all DNA profiles from unsolved crime scenes have been matched to individuals on the database.

Approximately 31% of all DNA profiles from unsolved crime scenes have been matched to other unsolved crime scenes on the database.

Approximately 78% of matches made between crime scenes and individuals have been in relation to burglaries.

A number of serious serial offenders have been identified and apprehended as a result of the database.

United States:
The FBI's Combined DNA Index System (CODIS) commenced operations in October 1998.

All states, along with the Federal Government, have implemented laws to collect DNA profiles from convicted offenders and place them on a database.

CODIS has three hierarchical tiers - local, state and national.

CODIS is installed in 137 forensic laboratories in 47 states and the District of Columbia.

To date, CODIS has assisted in over 1,900 investigations in 31 states.

The national tier of CODIS holds over 600,000 convicted offender DNA profiles, along with 26,000 crime scene samples.

Victorian DNA Database match information:
At 20 June 2002, the Victorian (ie VFSC) DNA database had detected the following previously unknown DNA profile matches:

<table>
<thead>
<tr>
<th>DNA database matches detected</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Convicted Offender to Crime</td>
<td>574</td>
</tr>
<tr>
<td>Suspect to Crime</td>
<td>148</td>
</tr>
<tr>
<td>Crime to Crime</td>
<td>353</td>
</tr>
<tr>
<td>TOTAL</td>
<td>1075</td>
</tr>
</tbody>
</table>

Table 5: Victorian DNA Database matches detected
The breakdown of DNA database matches by offence type to 20 June 2002 can be seen in table 6.

<table>
<thead>
<tr>
<th>Sample Type</th>
<th>Offence Type</th>
<th>Number of Matches</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Convicted Offender to crime</td>
<td>Burglary</td>
<td>431</td>
<td>574</td>
</tr>
<tr>
<td></td>
<td>Theft Of Motor Car</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Murder</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rape/Attempted Rape</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Armed Robbery</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Assault</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Drugs Traffic</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Incest</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aggravated Burglary</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Theft</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wilful and Obscene Exposure</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cause Serious Bodily Injury</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Firearms Offence</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Robbery</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Abduction</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Attempted Murder</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sexual Penetration</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Police Enquiry</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aggravated Rape</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stabbing</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hit Run</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Attempted Burglary</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Drugs</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Suspects to Crime</td>
<td>Suspicious Death</td>
<td>2</td>
<td>148</td>
</tr>
<tr>
<td></td>
<td>Burglary</td>
<td>98</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Armed Robbery</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sexual Assault</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rape</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Assault</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Missing Person</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Murder</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Abduction</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aggravated Burglary</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Theft</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Theft of Motor Car</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Police Enquiry</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Affray</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cause Serious Bodily Injury</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Crime to crime</td>
<td>Various Combinations (eg: Burglary with Burglary; Burglary with Rape)</td>
<td>353</td>
<td></td>
</tr>
</tbody>
</table>

**Table 6: Breakdown of Victorian DNA database matches by offence type**

The method by which DNA database matches is counted is currently under discussion by Australian forensic laboratories. It is essential that all laboratories count DNA database matches in the same way so that direct comparisons between databases can be made if required. This is currently not the situation.
A very simple example of how database matches can be counted differently is where person A matches crime sample B. The VFSC would count this as one match. Some other laboratories would count this as two matches, since A matches B, but B also matches A. In the latter instance, this laboratory would seemingly have detected twice the number of matches as the VFSC.

The VFSC currently reports DNA database matches to the Victoria Police Crime Department Forensic Procedures Implementation Team (FPIT) who then through further preliminary investigation, determine whether a particular reported match should be investigated in detail.

FPIT will formally cease to exist on 30 June 2002. A ‘DNA Unit’ will replace FPIT on 1 August 2002. This group will continue to perform the functions of FPIT in addition to other tasks.

2.1.3 Interpol
An Interpol European Working Party on DNA profiling was originally established in 1996 to provide a forum where European experts on DNA profiling could meet to set up guidelines and recommendations with a view to promote the wide use of a standard DNA profiling technique in Europe.

Based on experience from the countries already routinely using DNA profiling in their criminal investigations and taking into consideration the work already done in this field by other bodies such as the European Network of Forensic Science Institutes (ENFSI), the European DNA Profiling Group (EDNAP), the European Union Working Group on Police Cooperation and others, the Interpol European Working Party on DNA Profiling dealt with the following aspects of DNA profiling:

- Technical and scientific requirements (DNA technology)
- Principles for DNA sampling and evidence collection
- DNA databases
- Categories of offenders
- Quality control and accreditation
- Legal aspects
- Promotion and marketing

The final report of the Interpol European Working Party on DNA Profiling was endorsed by the 27th European Regional Conference in Dubrovnik (Croatia) 13-15 May 1998 and then submitted to the 67th General Assembly Session (Cairo, 22-27 October 1998). The working party was instructed to prepare guidelines and recommendations with a view to promoting the wider use of a standard DNA profiling technique in Europe as a powerful tool in criminal investigation (74)

The Interpol DNA Monitoring Expert Group (Interpol DNA MEG) is the successor to the Interpol European Working Party on DNA Profiling. Australia as an Interpol member, actively participates in the Interpol DNA MEG.

Interpol is proposing to establish an international DNA database for use by its Member States. Countries will be able to add profiles from their national or regional databases and compare their profiles with those supplied by participating Interpol Member States.
The potential benefits of a worldwide network of DNA databases are obvious. However, achieving this goal may take time given the various legislative controls present in different countries and jurisdictions.

2.2 The use of a DNA database

2.2.1 DNA database searches and the meaning of matches obtained.

When samples are DNA profiled it is desirable that a full profile result is obtained in the testing system used. The Victoria Forensic Science Centre (VFSC) uses the Profiler Plus™ system. When a full profile is obtained, 20 ‘results’ (ie allele designations) are obtained per sample. The alleles are given numerical names for Profiler Plus™ loci other than Amelogenin. The names are the letters of the alphabet for the locus Amelogenin. The numbers can be the same or different for different DNA sites. Note that there is a complicated scientific explanation for this naming convention (ie nomenclature), however, this explanation is not necessary for the purposes of this section.

A DNA profile therefore, is simply a series of numbers or letters that details the genotype of the person or sample that has been tested.

On occasion, a sample will not produce a full profile. It produces what is known as a ‘partial profile’. There can be many reasons why this occurs, but principally this is a feature of the quality and the quantity of the sample.

When a partial profile is obtained, less than 20 results or allele designations are obtained per sample.

In performing a DNA profile match search on a database, the best type of match is one where all 20 alleles from one sample match (ie are the same as) all 20 alleles from another sample.

This would be given a (20,0,0) designation using VFSC nomenclature, indicating that 20 matches, 0 mismatches and 0 null matches were obtained for the comparison.

Sometimes for example, a result of (19,0,1) may be obtained. This would indicate that 19 alleles matched between the two samples compared, there were no mismatches, and 1 result did not provide any information (ie it was a null result). A null result occurs when in one sample, one or more than one of the expected 20 alleles is not obtained (ie a partial profile only could be obtained on DNA testing).

This type of match between two profiles would still be considered as a match.

If however, a (19,1,0) result was obtained, this would indicate that the two samples in fact were not matching samples. Even though 19 alleles matched, one did not, which means that the samples could not have come from the same person or source. This result is not a match.
The construction of the VFSC DNA database allows database searches to be performed over any criteria from looking for matches of type (20,0,0) to any combination of (matches, mismatches, nulls) where the sum of matches plus mismatches plus nulls equals 20.

This flexibility allows the VFSC to search for possible matches of partial profiles to other partial or full/complete profiles. Partial profiles can contain any number of alleles from 19 to 1, therefore the possible setting of (match, mismatch, null) levels for searching the database is extremely varied.

The greater the number of matches (and therefore the lesser the number of nulls) scored for a particular database match (which may typically be between a crime scene sample and a person sample), the smaller the likelihood that a second person chosen at random from the Victorian population could have been the donor of the crime scene sample, if the person on the database who matched the crime scene sample was in fact not the true donor.

It is estimated that the most common Profiler Plus™ profile in the Victorian Caucasian population occurs in approximately 1 person in 98 million. This is a figure calculated for full profiles where 20 alleles have been obtained.

However, for partial profiles this frequency may be less (ie 1 in a figure less than 98 million eg 1 in 50 million).

The reporting of a match never proves that a particular person is the donor of the crime scene sample. The reporting of a match only indicates that the person could be the donor, and subsequent calculations (the likelihood ratio-explained later in this section) may be required to determine the relative value of that match.

Whilst routine VFSC match searches are performed from (20,0,0) to (16,0,0) (high stringency searches), on occasion it is necessary to perform matches at a lower level (low stringency searches). This is usually prompted by trying to see if any full or partial DNA profiles in the database match any other partial DNA profile on the database.

Table 7 shows an example of some database match results:
Forensic sampling and DNA databases: background/issues paper

<table>
<thead>
<tr>
<th>Sample</th>
<th>DNA test area A</th>
<th>DNA test area B</th>
<th>DNA test area C</th>
<th>DNA test area D</th>
<th>DNA test area E</th>
<th>DNA test area F</th>
<th>DNA test area G</th>
<th>DNA test area H</th>
<th>DNA test area I</th>
<th>DNA test area J</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partial DNA profile from crime scene (M)</td>
<td>1,2</td>
<td>3,4</td>
<td>5,6</td>
<td>7,8</td>
<td>No result</td>
<td>No result</td>
<td>No result</td>
<td>No result</td>
<td>No result</td>
<td>No result</td>
</tr>
<tr>
<td>Person N</td>
<td>1,2</td>
<td>3,4</td>
<td>5,6</td>
<td>7,8</td>
<td>9,10</td>
<td>11,12</td>
<td>13,14</td>
<td>15,16</td>
<td>17,18</td>
<td>X,Y*</td>
</tr>
<tr>
<td>Person L</td>
<td>1,2</td>
<td>3,4</td>
<td>5,6</td>
<td>7,8</td>
<td>21,22</td>
<td>23,24</td>
<td>25,26</td>
<td>27,28</td>
<td>29,30</td>
<td>X,X**</td>
</tr>
<tr>
<td>Partial DNA profile from another crime scene (P)</td>
<td>1,2</td>
<td>3,4</td>
<td>5,6</td>
<td>7,8</td>
<td>21,22</td>
<td>23,24</td>
<td>No result</td>
<td>No result</td>
<td>No result</td>
<td>No result</td>
</tr>
</tbody>
</table>

* this result indicates that the sample has come from a male
** this result indicates that the sample has probably come from a female.

Table 7: DNA profile results

If a search of the database was performed at (6,4,10) the result table above could be obtained as an example. Note that M is the profile being searched against the rest of the DNA database.

It can be observed that N, L and P all match M. This is clear for DNA test areas A-D. DNA test areas E-J in profile M gave no results, therefore it is possible that person N could be the same DNA type as M (if results for M could be obtained). Equally it is possible that person L could be the same DNA type as M (if results for M could be obtained). Therefore both N and L could match M based on the available information.

However, N and L do not match each other as DNA test areas E-J show that these individuals have different DNA types at these sites. Sample P matches L but does not match N.

It is important to realise that any likelihood ratio number greater than 1 is evidence suggesting that the proposed donor of the sample is the true donor of the sample. (Note that likelihood ratios are explained later in this section.) If calculations to determine the evidential value of these matches were performed the following might be obtained as an example (refer to table 8):
Forensic sampling and DNA databases: background/issues paper

<table>
<thead>
<tr>
<th>Match between</th>
<th>Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>M and N</td>
<td>5000</td>
</tr>
<tr>
<td>M and L</td>
<td>5000</td>
</tr>
<tr>
<td>N and L</td>
<td>0</td>
</tr>
<tr>
<td>M and P</td>
<td>5000</td>
</tr>
<tr>
<td>L and P</td>
<td>25000</td>
</tr>
<tr>
<td>P and N</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 8: Likelihood ratios for matches between profiles

The value of the match between M and N, and M and L are the same as the only DNA sites that contribute to the calculation of the likelihood ratio are sites A-D. Even though profiles N and L both match M, because profile M has no results at DNA test sites E-J, it is not possible to determine if profile N or L would match if those sites had yielded results.

This shows that it is possible to have more than one person appear to match a sample in the database. The fact that this can occur does not mean that the database is not working, or that the result is nonsensical or meaningless. It demonstrates that the database is working and that the level of information that has been obtained is less discriminating than may be obtained when results from more DNA loci are obtained in a profile.

An even simpler example is the following. A crime is committed. The victim says the offender was a man in a suit. This means that all men in suits could be the possible offender. (*)This includes men with long hair, short hair, and bald men. If through some other means it was known that the offender was bald, then the men with short hair and long hair, although they were initially included, can now be excluded from the list of possible offenders. (**)

(*) is the result off the DNA database as illustrated in the example above
(**) is subsequent investigative work performed by Police once they have received the list of possible suspects found from the DNA database search.

Matches to partial profiles may be less informative, but can provide a small pool of possible suspects and/or other cases that can then be investigated further by the Police in efforts to find the true offender. These types of match searches, whilst potentially being less discriminating, can provide Police with further avenues for investigation particularly in cases where conventional police investigative techniques have not identified an offender, or have identified a number of individuals amongst which may be the true offender.

Likelihood ratio:
In determining the value of the DNA profile matches detected for database searches for the VFSC DNA database, a ‘likelihood ratio’ is calculated.
The likelihood ratio is the figure obtained in consideration of two alternative hypotheses:
A that the DNA in the sample (in this case, this is the DNA profile being compared to the rest of the DNA database) originated from the donor of the DNA in the matching DNA profile, or
B that the DNA in the sample originated from another person chosen at random from the Victorian Caucasian population

The match is the calculated number of times more likely if the DNA in the sample originated from the donor of the DNA in the matching DNA profile than if it originated from another person chosen at random from the Victorian Caucasian population. This figure is called the likelihood ratio.

Note however, that the number of people on the database at the time of the database search limits the number of people possibly matching the DNA from a crime sample. If more people were on the database, there may be potentially more individuals who could also match the DNA on a particular crime stain.

A database search can however provide the Police with a limited pool of potential suspects to a crime. The Police can now further investigate these individuals to determine through other investigative techniques whether any of these individuals were perhaps the offender(s) for that crime. It may transpire that the individual(s) identified as being possible contributors of the DNA on the crime sample were not the true offenders, and that some other person is, and that that person is not on the DNA database.

2.2.2 What profiles should be put on a DNA database-single source DNA samples and/or DNA mixture profiles?

From section 2.2.1 it can be seen that there are various stringencies that can be applied to DNA database match searches for profiles from single DNA source samples. Performing matches of this type requires considerable computing power, increasing as the number of profiles on the database increases.

The complexity of match searches for DNA profiles that contain DNA from more than one source (ie mixtures of DNA) is even greater. The following example demonstrates what needs to be considered for a database match search of a mixture compared to that of a sample containing a single source of DNA.

Example:
A crime sample (single source) contains alleles 10 and 11 at a particular DNA locus. DNA database person samples must now be compared to this crime sample. The comparison occurs in the following way:

Crime sample alleles 10 11

Database sample alleles A B (where A and B represent the alleles present in each particular person sample on the database.)
The arrows represent the comparisons that are made. For example, allele A is compared to allele 10 and 11 to see if allele A is the same or different to 10 and 11. Allele B is compared to allele 10 and 11 to see if allele B is the same or different to allele 10 and 11.

If A and B in this example were alleles 10 and 11, a perfect match would be obtained. If alleles A and B were alleles 9 and 10, a partial match between the crime and person sample would be obtained.

If A and B in this example were alleles 8 and 9, then no match would be obtained.

Similar comparisons occur at each of the other DNA loci in the DNA profile.

Now consider a crime sample that contains a mixture of DNA from more than one individual. The crime sample contains the alleles 10, 11, 14, 17 at locus one, and alleles 5, 6, 7 at locus two. Assume that all alleles are present at approximately the same amounts at locus 1. At locus 2 allele 7 is present at twice the amount of alleles 5 and 6. DNA database person samples are compared to this crime sample to attempt to determine possible donors to the crime sample. Alleles A and B represent the alleles present in each particular person sample on the database. The comparison must occur in the following way:

<table>
<thead>
<tr>
<th>Crime sample alleles at locus 1</th>
<th>10</th>
<th>11</th>
<th>14</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database sample alleles at locus 1</td>
<td>A</td>
<td>B</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

At locus 2, the following comparisons are being made:

<table>
<thead>
<tr>
<th>Crime sample alleles at locus 2</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database sample alleles at locus 2</td>
<td>C</td>
<td>D</td>
<td></td>
</tr>
</tbody>
</table>

The set of comparisons for locus 2 must occur after each set of comparisons for locus 1. For example alleles A and B are compared to alleles 10 and 11. Simultaneously, alleles C and D are compared to alleles 5 and 6. Alleles A and B are then compared to alleles 10 and 14. Alleles C and D are held constant and still are compared to alleles 5 and 6. Alleles A and B are compared then to alleles 10 and 17 and alleles C and D are still held constant and compared to alleles 5 and 6. When all the options for locus 1 are exhausted, variations begin with locus 2 whilst locus 1 is held constant. For example, alleles A and B are compared to alleles 10 and 11, and simultaneously alleles C and D are compared to alleles 6 and 7. This example illustrates that the total number and types of comparisons made for a sample of DNA from a mixture is considerably greater than for a DNA profile obtained from a single source.
It can easily be seen that when ten loci are being considered that the number of combinations of searches for matches of reference samples to the DNA mixture sample becomes considerable.

This complexity increases again when DNA mixture samples contain unequal proportions of DNA from each potential contributor. When the amount of DNA from one contributor is very minor, it is possible that poor amplification of that component may occur simply because there is insufficient starting DNA. This is known as a stochastic effect. When this occurs it is possible that not all alleles that are truly present in a person’s DNA may be amplified. The amplified product DNA produced may then not completely match the DNA that is that person’s true genotype. In this case it can become difficult to determine conclusively if a person could be a potential contributor to the mixture of DNA in the sample.

Ironically, the complexity of resolution of mixtures can sometimes be reduced when the amount of DNA from the various contributors to the sample is grossly disproportionate. With this type of mixture it can be possible sometimes to separate the individual major and minor components of the DNA. When this is possible, the major and/or minor DNA profile can be individually added to a DNA database, and match runs performed looking at each individual component.

The complexity of resolution of possible contributors to a DNA mixture through a database search also increases dramatically the greater the number of potential contributors to the mixture.

For these reasons, DNA database searches are usually only performed on single DNA source samples, or on mixture samples where the major and/or minor components can be easily resolved. Creating computer algorithms to enable database searches for DNA mixtures is extremely complex. Considerable computing skill and power would be required.

Therefore DNA database searches are currently limited to single DNA source samples. It is very unlikely that in the foreseeable future that this will change.

2.2.3 Frequency of DNA database match runs
The frequency with which DNA database match runs are performed is to some extent purely a subjective decision. Ideally it makes sense to perform a match run whenever new profiles are added to the database, and this is the intended mode of operation of the Australian National Criminal Investigation DNA Database. This will ensure that new matches are detected as rapidly as possible.

Some laboratories operate by performing match runs at a specified time of the day, some perform match runs daily when new profiles have been added the previous day. The decision as to how often match runs are performed is often the culmination of consideration of issues such as police and forensic laboratory resourcing, the structure and capabilities of the forensic laboratory’s DNA database, service level agreements that may be in place, legislation and so on.
### 2.2.4 Permissible matching of DNA profiles

In Victoria, s464ZGI of the Crimes (DNA Database) Act 2002 (54) details the permissible matching of DNA profiles. Table 9 below shows match searches that are allowable.

<table>
<thead>
<tr>
<th>Profile to be matched</th>
<th>Is matching permitted?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Column 1</td>
</tr>
<tr>
<td>Crime scene</td>
<td>yes</td>
</tr>
<tr>
<td>Suspects</td>
<td>yes</td>
</tr>
<tr>
<td>Volunteers (limited purposes)</td>
<td>yes</td>
</tr>
<tr>
<td>Volunteers (unlimited purposes)</td>
<td>no</td>
</tr>
<tr>
<td>Serious offenders</td>
<td>yes</td>
</tr>
<tr>
<td>Missing persons</td>
<td>yes</td>
</tr>
<tr>
<td>Unknown deceased persons</td>
<td>yes</td>
</tr>
</tbody>
</table>

**Table 9: Permissible matching of DNA profiles under s464ZGI of the Crimes (DNA Database) Act 2002**

Most Australian States and Territories have similar tables in their legislation, with the current exception of the Northern Territory.

### 2.2.5 Destruction of information

S464 of the Victorian Crimes Act (53) contains specific information about the length of time under which samples taken under this legislation can be used.

S464O refers to the destruction of fingerprint records.

S464ZFC refers to the destruction of information relating to DNA samples where an order for the retention of a sample, and any related material and information following finding of guilt has not been obtained within 6 months after the expiry of the appeal period in respect of the offence, or the final determination of an appeal (whichever is the latter).

S464ZG refers to the destruction of samples, and any related material and information from a forensic procedure conducted on a person and where the person has not been charged with a relevant offence at the end of a period of 12 months after the conduct of the procedure, or if the person has been so charged but the charge is not proceeded with or the person is not found guilty of the offence or any other relevant offence, whether on appeal or otherwise, before the end of that period. A sample and any related material and information must be destroyed immediately after that period of 12 months or, within 1 month after the conclusion of the proceedings and the end of any appeal period, or if the proceedings have been adjourned under s75 of the Sentencing Act 1991 (75) within 1 month of dismissal under that section. A member
of the police force, before the end of the periods specified in the previous sentence apply to the Magistrate’s Court or the Children’s Court for an order extending the period. If the Court makes such an order, it must give reasons for it’s decision and cause a copy of the order to be served on the person on whom the forensic procedure was conducted.

These legislative requirements determine time frames under which information from a person may be used. The consequence is that there is usually a complicated support system around the operation of a DNA database, ensuring that samples and information relating to samples are destroyed as required, and that DNA database match runs are not performed on information from individuals inappropriately.

Other states and territories have their own legislative requirements in this regard to sample and information destruction.
3 References


3 Takayama (1912) Japanese J Toxicology

4 Kerr D, Mason V (1926) Brit Med J 1, 134-136


13 Kaye S (1949) Acid phosphatase test for the identification of seminal stains. J Lab Clin Med 34, 728-732

14 Lindquist F (1946) Nature 158, 710


21 Biology Methods Manual, Metropolitan Police Forensic Science Laboratory (1978)


31 Lareu M, Rodriguez-Calvo M, Barros F, Carracedo A (1992). The usefulness of chelating resins for DNA extraction from forensic material prior to PCR amplification. Advances in Forensic Haemogenetics 4, 47-49


Singer-Sam J, Tanguay R. “Use of chelex to improve PCR signal from a small number of cells” in Amplifications, A Forum for PCR Users, Sept 1989, Issue 3


Sharma, V., and Litt, M. 1992 “Tetrancleotide repeat polymorphism at the D21S11 locus.” Human Molecular Genetics 1, 67


53 Crimes Act 1958

54 Crimes (DNA Database) Act 2002


(RFLP) database for Punjabis in East Punjab, India. For Sci Int 79, 187-198


63 Morling N (1997) DNA in forensic genetics in Denmark. For Sci Int 88, 43-46


65 Johnsson V (1997) DNA testing in crime investigation in Finland. For Sci Int 88, 51-52


75 Sentencing Act 1991
4 glossary

portions of the terms were taken directly or modified from definitions in the u.s. congress office of technology assessment document mapping our genes the genome projects: how big, how fast? ota-ba-373, washington, d.c.: u.s. government printing office, april 1988. the original source is at http://www.ornl.gov/techresources/human_genome/publicat/primer/glossary.html see also, http://www.nhgri.nih.gov/dir/vip/glossary/

portions of the terms were also taken from a glossary compiled by the forensic science centre, adelaide.

the glossary contains terms that are found in the text of this document. it also contains terms which do not appear in this document, but which may be useful for future reference.

a

adenine (a): a nitrogenous base, one member of the base pair a-t (adenine-thymine).

alleles: alternative forms of a genetic locus; a single allele for each locus is inherited separately from each parent (e.g., at a locus for eye colour the allele might result in blue or brown eyes).

amino acid: any of a class of 20 molecules that are combined to form proteins in living things. the sequence of amino acids in a protein and hence protein function are determined by the genetic code.

amplification: an increase in the number of copies of a specific dna fragment; can be in vivo or in vitro. see cloning, polymerase chain reaction.

arrayed library: individual primary recombinant clones (hosted in phage, cosmid, yac, or other vector) that are placed in two-dimensional arrays in microtiter dishes. each primary clone can be identified by the identity of the plate and the clone location (row and column) on that plate. arrayed libraries of clones can be used for many applications, including screening for a specific gene or genomic region of interest as well as for physical mapping. information gathered on individual clones from various genetic linkage and physical map analyses is entered into a relational database and used to construct physical and genetic linkage maps simultaneously; clone identifiers serve to interrelate the multilevel maps. compare library, genomic library.

autoradiography: a technique that uses x-ray film to visualize radioactively labelled molecules or fragments of molecules; used in analyzing length and number of dna fragments after they are separated by gel electrophoresis.

autosome: a chromosome not involved in sex determination. the diploid human genome consists of 46 chromosomes, 22 pairs of autosomes, and 1 pair of sex chromosomes (the x and y chromosomes).

b

bacteriophage: see phage.


**Base pair (bp):** Two nitrogenous bases (*adenine* and *thymine* or *guanine* and *cytosine*) held together by weak bonds. Two strands of DNA are held together in the shape of a double helix by the bonds between base pairs.

**Base sequence:** The order of *nucleotide* bases in a DNA molecule.

**Base sequence analysis:** A method, sometimes automated, for determining the *base sequence*.

**Biotechnology:** A set of biological techniques developed through basic research and now applied to research and product development. In particular, the use by industry of *recombinant DNA*, cell fusion, and new bioprocessing techniques.

**bp:** See *base pair*.

---

**cDNA:** See *complementary DNA*.

**Centimorgan (cM):** A unit of measure of *recombination* frequency. One centimorgan is equal to a 1% chance that a marker at one genetic *locus* will be separated from a marker at a second locus due to *crossing over* in a single generation. In human beings, 1 centimorgan is equivalent, on average, to 1 million *base pairs*.

**Centromere:** A specialized *chromosome* region to which spindle fibres attach during cell division.

**Chromosomes:** The self-replicating genetic structures of cells containing the cellular DNA that bears in its *nucleotide* sequence the linear array of *genes*. In *prokaryotes*, chromosomal DNA is circular, and the entire genome is carried on one chromosome. *Eukaryotic* genomes consist of a number of chromosomes whose DNA is associated with different kinds of *proteins*.

**Clone bank:** See *genomic library*.

**Clones:** A group of cells derived from a single ancestor.

**Control samples:** Samples used to monitor the success of various stages of the testing process. Reagent blanks (or extraction negatives) and negative controls (or amplification negatives) do not contain DNA and test for contamination. Positive controls contain DNA and test the amplification and DNA fragment separation steps.

**Cloning:** The process of asexually producing a group of cells (clones), all genetically identical, from a single ancestor. In *recombinant DNA* technology, the use of DNA manipulation procedures to produce multiple copies of a single *gene* or segment of DNA is referred to as cloning DNA.

**Cloning vector:** DNA molecule originating from a *virus*, a *plasmid*, or the cell of a higher organism into which another DNA fragment of appropriate size can be integrated without loss of the vectors capacity for self-replication; vectors introduce foreign DNA into host cells, where it can be reproduced in large quantities. Examples are *plasmids*, *cosmids*, and *yeast artificial chromosomes*; vectors are often *recombinant* molecules containing DNA sequences from several sources.

**cM:** See cM.

**Code:** See *genetic code*.

**Codon:** See *genetic code*.

**Complementary DNA (cDNA):** DNA that is synthesized from a *messenger RNA* template; the single-stranded form is often used as a *probe* in *physical mapping*.

**Complementary sequences:** *Nucleic acid* base sequences that can form a double-stranded structure by matching *base pairs*; the complementary sequence to G- T- A- C is C- A- T- G.

**Conserved sequence:** A *base sequence* in a DNA molecule (or an *amino acid* sequence in a *protein*) that has remained essentially unchanged throughout evolution.
Contig map: A map depicting the relative order of a linked library of small overlapping clones representing a complete chromosomal segment.

Contigs: Groups of clones representing overlapping regions of a genome.

Cosmid: Artificially constructed cloning vector containing the cos gene of phage lambda. Cosmids can be packaged in lambda phage particles for infection into E. coli; this permits cloning of larger DNA fragments (up to 45 kb) than can be introduced into bacterial hosts in plasmid vectors.

Crossing over: The breaking during meiosis of one maternal and one paternal chromosome, the exchange of corresponding sections of DNA, and the rejoining of the chromosomes. This process can result in an exchange of alleles between chromosomes. Compare recombination.

Cytosine (C): A nitrogenous base, one member of the base pair G-C (guanine and cytosine).

Deoxyribonucleotide: See nucleotide.

Denature: to treat (a protein etc) by chemical or physical means, such as adding acid or heating, so as to cause loss of solubility, biological activity etc. To change the structure of a protein or other substance, and thus change its properties as by adding heat, pressure or chemicals.

Diploid: A full set of genetic material, consisting of paired chromosomes one chromosome from each parental set. Most animal cells except the gametes have a diploid set of chromosomes. The diploid human genome has 46 chromosomes. Compare haploid.

DNA (deoxyribonucleic acid): The molecule that encodes genetic information. DNA is a double- base pairs of nucleotides. The four nucleotides in DNA contain the bases constructed as a double stranded molecule held together by weak bonds between base pairs of nucleotides. The four nucleotides in DNA contain the bases: adenine (A), guanine (G), cytosine (C), and thymine (T). In nature, base pairs form only between A and T and between G and C; thus the base sequence of each single strand can be deduced from that of its partner.

DNA probes: See probe.

DNA replication: The use of existing DNA as a template for the synthesis of new DNA strands. In humans and other eukaryotes, replication occurs in the cell nucleus.

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

Domain: A discrete portion of a protein with its own function. The combination of domains in a single protein determines its overall function.

Double helix: The shape that two linear strands of DNA assume when bonded together.

E. coli: Common bacterium that has been studied intensively by geneticists because of its small genome size, normal lack of pathogenicity, and ease of growth in the laboratory.

Electrophoresis: A method of separating large molecules (such as DNA fragments or proteins) from a mixture of similar molecules. An electric current is passed through a medium containing the mixture, and each kind of molecule travels through the medium at a different rate, depending on its electrical charge and size. Separation is
based on these differences. Agarose and acrylamide gels are the media commonly used for electrophoresis of proteins and nucleic acids.

**Endonuclease:** An enzyme that cleaves its nucleic acid substrate at internal sites in the nucleotide sequence.

**Enzyme:** A protein that acts as a catalyst, speeding the rate at which a biochemical reaction proceeds but not altering the direction or nature of the reaction.

**EST:** Expressed sequence tag. See sequence tagged site.

**Eukaryote:** Cell or organism with membrane-bound, structurally discrete nucleus and other well-developed subcellular compartments. Eukaryotes include all organisms except viruses, bacteria, and blue-green algae. Compare prokaryote. See chromosomes.

**Evolutionarily conserved:** See conserved sequence.

**Exogenous DNA:** DNA originating outside an organism.

**Exons:** The protein-coding DNA sequences of a gene. Compare introns.

**Exonuclease:** An enzyme that cleaves nucleotides sequentially from free ends of a linear nucleic acid substrate.

**Expressed gene:** See gene expression.

**F**

**FISH (fluorescence in situ hybridization):** A physical mapping approach that uses fluorescein tags to detect hybridization of probes with metaphase chromosomes and with the less-condensed somatic interphase chromatin.

**Flow cytometry:** Analysis of biological material by detection of the light-absorbing or fluorescing properties of cells or subcellular fractions (i.e., chromosomes) passing in a narrow stream through a laser beam. An absorbance or fluorescence profile of the sample is produced. Automated sorting devices, used to fractionate samples, sort successive droplets of the analyzed stream into different fractions depending on the fluorescence emitted by each droplet.

**Flow karyotyping:** Use of flow cytometry to analyze and/or separate chromosomes on the basis of their DNA content.

**Frequency:** The proportion or occurrence of an allele, genotype or profile in a population.

**F_{st}:** Measure of coancestry or relatedness of alleles. The probability that alleles within a population have a common ancestor

**G**

**Gamete:** Mature male or female reproductive cell (sperm or ovum) with a haploid set of chromosomes (23 for humans).

**Gene:** The fundamental physical and functional unit of heredity. A gene is an ordered sequence of nucleotides located in a particular position on a particular chromosome that encodes a specific functional product (i.e., a protein or RNA molecule). See gene expression.

**Gene expression:** The process by which a genes coded information is converted into the structures present and operating in the cell. Expressed genes include those that are transcribed into mRNA and then translated into protein and those that are transcribed into RNA but not translated into protein (e.g., transfer and ribosomal RNAs).

**Gene families:** Groups of closely related genes that make similar products.

**Gene library:** See genomic library.
Gene mapping: Determination of the relative positions of genes on a DNA molecule (chromosome or plasmid) and of the distance, in linkage units or physical units, between them.

Gene product: The biochemical material, either RNA or protein, resulting from expression of a gene. The amount of gene product is used to measure how active a gene is; abnormal amounts can be correlated with disease-causing alleles.

Genetic code: The sequence of nucleotides, coded in triplets (codons) along the mRNA, that determines the sequence of amino acids in protein synthesis. The DNA sequence of a gene can be used to predict the mRNA sequence, and the genetic code can in turn be used to predict the amino acid sequence.

Genetic engineering technologies: See recombinant DNA technologies.

Genetic map: See linkage map.

Genetic material: See genome.

Genetics: The study of the patterns of inheritance of specific traits.

Genome: All the genetic material in the chromosomes of a particular organism; its size is generally given as its total number of base pairs.

Genome projects: Research and technology development efforts aimed at mapping and sequencing some or all of the genome of human beings and other organisms.

Genomic library: A collection of clones made from a set of randomly generated overlapping DNA fragments representing the entire genome of an organism. Compare library, arrayed library.

Genotype: The combination of the two alleles at a locus

Guanine (G): A nitrogenous base, one member of the base pair G-C (guanine and cytosine).

H

Haploid: A single set of chromosomes (half the full set of genetic material), present in the egg and sperm cells of animals and in the egg and pollen cells of plants. Human beings have 23 chromosomes in their reproductive cells. Compare diploid.

Heterozygosity: The presence of different alleles at one or more loci on homologous chromosomes.

Homeobox: A short stretch of nucleotides whose base sequence is virtually identical in all the genes that contain it. It has been found in many organisms from fruit flies to human beings. In the fruit fly, a homeobox appears to determine when particular groups of genes are expressed during development.

Homologies: Similarities in DNA or protein sequences between individuals of the same species or among different species.

Homologous chromosomes: A pair of chromosomes containing the same linear gene sequences, each derived from one parent.

Homozygote: Where the two alleles at a locus are the same. Each parent contributes the same allele to the child.

Human gene therapy: Insertion of normal DNA directly into cells to correct a genetic defect.

Human Genome Initiative: Collective name for several projects begun in 1986 by DOE to (1) create an ordered set of DNA segments from known chromosomal locations, (2) develop new computational methods for analyzing genetic map and DNA sequence data, and (3) develop new techniques and instruments for detecting and analyzing DNA. This DOE initiative is now known as the Human Genome Program. The national effort, led by DOE and NIH, is known as the Human Genome Project.
Hybridization: The process of joining two complementary strands of DNA or one each of DNA and RNA to form a double-stranded molecule.

Informatics: The study of the application of computer and statistical techniques to the management of information. In genome projects, informatics includes the development of methods to search databases quickly, to analyze DNA sequence information, and to predict protein sequence and structure from DNA sequence data.

In situ hybridization: Use of a DNA or RNA probe to detect the presence of the complementary DNA sequence in cloned bacterial or cultured eukaryotic cells.

Interphase: The period in the cell cycle when DNA is replicated in the nucleus; followed by mitosis.

Introns: The DNA base sequences interrupting the protein-coding sequences of a gene; these sequences are transcribed into RNA but are cut out of the message before it is translated into protein. Compare exons.

In vitro: Outside a living organism.

Karyotype: A photomicrograph of an individual's chromosomes arranged in a standard format showing the number, size, and shape of each chromosome type; used in low-resolution physical mapping to correlate gross chromosomal abnormalities with the characteristics of specific diseases.

Kb: See kilobase.

Kilobase (kb): Unit of length for DNA fragments equal to 1000 nucleotides.

Library: An unordered collection of clones (i.e., cloned DNA from a particular organism), whose relationship to each other can be established by physical mapping. Compare genomic library, arrayed library.

Likelihood ratio: A ratio comparing the probability (or likelihood) of a crime stain and a reference sample having matching profiles given competing scenarios. E.g., prosecution hypothesis that the crime stain was left by the defendant compared to the defence hypothesis that the crime stain was left by an unknown person. LR’s greater than one support the prosecution scenario while LR’s less than one support the defence scenario.

Linkage: The proximity of two or more markers (e.g., genes, RFLP markers) on a chromosome; the closer together the markers are, the lower the probability that they will be separated during DNA repair or replication processes (binary fission in prokaryotes, mitosis or meiosis in eukaryotes), and hence the greater the probability that they will be inherited together.

Linkage map: A map of the relative positions of genetic loci on a chromosome, determined on the basis of how often the loci are inherited together. Distance is measured in centimorgans (cM).

Localyze: Determination of the original position (locus) of a gene or other marker on a chromosome.

Locus (pl. loci): The position on a chromosome of a gene or other chromosome marker; also, the DNA at that position. The use of locus is sometimes restricted to mean regions of DNA that are expressed. See gene expression.
**M**

**Macrorestriction map:** Map depicting the order of and distance between sites at which restriction enzymes cleave chromosomes.

**Mapping:** See gene mapping, linkage map, physical map.

**Match probability:** Estimates the probability of seeing a profile, given (or conditional on) it has already been seen. This is also known as a conditional probability statement.

**Marker:** An identifiable physical location on a chromosome (e.g., restriction enzyme cutting site, gene) whose inheritance can be monitored. Markers can be expressed regions of DNA (genes) or some segment of DNA with no known coding function but whose pattern of inheritance can be determined. See RFLP, restriction fragment length polymorphism.

**Mb:** See megabase.

**Megabase (Mb):** Unit of length for DNA fragments equal to 1 million nucleotides and roughly equal to 1 cM.

**Meiosis:** The process of two consecutive cell divisions in the diploid progenitors of sex cells. Meiosis results in four rather than two daughter cells, each with a haploid set of chromosomes.

**Messenger RNA (mRNA):** RNA that serves as a template for protein synthesis. See genetic code.

**Metaphase:** A stage in mitosis or meiosis during which the chromosomes are aligned along the equatorial plane of the cell.

**Mitosis:** The process of nuclear division in cells that produces daughter cells that are genetically identical to each other and to the parent cell.

**mRNA:** See messenger RNA.

**Multifactorial or multigenic disorders:** See polygenic disorders.

**Multiplexing:** A sequencing approach that uses several pooled samples simultaneously, greatly increasing sequencing speed.

**Multiplex PCR:** A number of loci are amplified simultaneously in the one reaction.

**Mutation:** Any heritable change in DNA sequence. Compare polymorphism.

**N**

**Nitrogenous base:** A nitrogen-containing molecule having the chemical properties of a base.

**Nucleic acid:** A large molecule composed of nucleotide subunits.

**Nucleotide:** A subunit of DNA or RNA consisting of a nitrogenous base (adenine, guanine, thymine, or cytosine in DNA; adenine, guanine, uracil, or cytosine in RNA), a phosphate molecule, and a sugar molecule (deoxyribose in DNA and ribose in RNA). Thousands of nucleotides are linked to form a DNA or RNA molecule. See DNA, base pair, RNA.

**Nucleus:** The cellular organelle in eukaryotes that contains the genetic material.

**O**

**Oncogene:** A gene, one or more forms of which is associated with cancer. Many oncogenes are involved, directly or indirectly, in controlling the rate of cell growth.

**Overlapping clones:** See genomic library.

**P**

**PCR:** See polymerase chain reaction.

**Phage:** A virus for which the natural host is a bacterial cell.
Physical map: A map of the locations of identifiable landmarks on DNA (e.g., restriction enzyme cutting sites, genes), regardless of inheritance. Distance is measured in base pairs. For the human genome, the lowest-resolution physical map is the banding patterns on the 24 different chromosomes; the highest-resolution map would be the complete nucleotide sequence of the chromosomes.

Plasmid: Autonomously replicating, extrachromosomal circular DNA molecules, distinct from the normal bacterial genome and nonessential for cell survival under nonselective conditions. Some plasmids are capable of integrating into the host genome. A number of artificially constructed plasmids are used as cloning vectors.

Polygenic disorders: Genetic disorders resulting from the combined action of alleles of more than one gene (e.g., heart disease, diabetes, and some cancers). Although such disorders are inherited, they depend on the simultaneous presence of several alleles; thus the hereditary patterns are usually more complex than those of single-gene disorders. Compare single-gene disorders.

Polymerase chain reaction (PCR): A method for amplifying a DNA base sequence using a heat-stable polymerase and two 20-base primers, one complementary to the (+)-strand at one end of the sequence to be amplified and the other complementary to the (-)-strand at the other end. Because the newly synthesized DNA strands can subsequently serve as additional templates for the same primer sequences, successive rounds of primer annealing, strand elongation, and dissociation produce rapid and highly specific amplification of the desired sequence. PCR also can be used to detect the existence of the defined sequence in a DNA sample.

Polymerase, DNA or RNA: Enzymes that catalyze the synthesis of nucleic acids on preexisting nucleic acid templates, assembling RNA from ribonucleotides or DNA from deoxyribonucleotides.

Polymorphism: Difference in DNA sequence among individuals. Genetic variations occurring in more than 1% of a population would be considered useful polymorphisms for genetic linkage analysis. Compare mutation.

Primer: Short preexisting polynucleotide chain to which new deoxyribonucleotides can be added by DNA polymerase.

Profile: The combination of genotypes from all the loci investigated.

Probe: Single-stranded DNA or RNA molecules of specific base sequence, labelled either radioactively or immunologically, that are used to detect the complementary base sequence by hybridization.

Prokaryote: Cell or organism lacking a membrane-bound, structurally discrete nucleus and other subcellular compartments. Bacteria are prokaryotes. Compare eukaryote. See chromosomes.

Promoter: A site on DNA to which RNA polymerase will bind and initiate transcription.

Protease: any of various enzymes that break down proteins into simpler compounds.

Protein: A large molecule composed of one or more chains of amino acids in a specific order; the order is determined by the base sequence of nucleotides in the gene coding for the protein. Proteins are required for the structure, function, and regulation of the body's cells, tissues, and organs, and each protein has unique functions. Examples are hormones, enzymes, and antibodies.

Purine: A nitrogen-containing, single-ring, basic compound that occurs in nucleic acids. The purines in DNA and RNA are adenine and guanine.

Pyrimidine: A nitrogen-containing, double-ring, basic compound that occurs in nucleic acids. The pyrimidines in DNA are cytosine and thymine; in RNA, cytosine and uracil.
Rare- cutter enzyme: See restriction enzyme cutting site.

Recombinant clones: Clones containing recombinant DNA molecules. See recombinant DNA technologies.

Recombinant DNA molecules: A combination of DNA molecules of different origin that are joined using recombinant DNA technologies.

Recombinant DNA technologies: Procedures used to join together DNA segments in a cell- free system (an environment outside a cell or organism). Under appropriate conditions, a recombinant DNA molecule can enter a cell and replicate there, either autonomously or after it has become integrated into a cellular chromosome.

Recombination: The process by which progeny derive a combination of genes different from that of either parent. In higher organisms, this can occur by crossing over.

Regulatory regions or sequences: A DNA base sequence that controls gene expression.

Resolution: Degree of molecular detail on a physical map of DNA, ranging from low to high.

Restriction enzyme, endonuclease: A protein that recognizes specific, short nucleotide sequences and cuts DNA at those sites. Bacteria contain over 400 such enzymes that recognize and cut over 100 different DNA sequences. See restriction enzyme cutting site.

Restriction enzyme cutting site: A specific nucleotide sequence of DNA at which a particular restriction enzyme cuts the DNA. Some sites occur frequently in DNA (e.g., every several hundred base pairs), others much less frequently (rare- cutter; e.g., every 10,000 base pairs).

Restriction fragment length polymorphism (RFLP): Variation between individuals in DNA fragment sizes cut by specific restriction enzymes; polymorphic sequences that result in RFLPs are used as markers on both physical maps and genetic linkage maps. RFLPs are usually caused by mutation at a cutting site. See marker.

RFLP: See restriction fragment length polymorphism.

Ribonucleic acid (RNA): A chemical found in the nucleus and cytoplasm of cells; it plays an important role in protein synthesis and other chemical activities of the cell. The structure of RNA is similar to that of DNA. There are several classes of RNA molecules, including messenger RNA, transfer RNA, ribosomal RNA, and other small RNAs, each serving a different purpose.

Ribonucleotides: See nucleotide.

Ribosomal RNA (rRNA): A class of RNA found in the ribosomes of cells.

Ribosomes: Small cellular components composed of specialized ribosomal RNA and protein; site of protein synthesis. See ribonucleic acid (RNA).

RNA: See ribonucleic acid.

Sequence: See base sequence.

Sequence tagged site (STS): Short (200 to 500 base pairs) DNA sequence that has a single occurrence in the human genome and whose location and base sequence are known. Detectable by polymerase chain reaction, STSs are useful for localizing and orienting the mapping and sequence data reported from many different laboratories and serve as landmarks on the developing physical map of the human genome. Expressed sequence tags (ESTs) are STSs derived from cDNAs.
Sequencing: Determination of the order of nucleotides (base sequences) in a DNA or RNA molecule or the order of amino acids in a protein.

Sex chromosomes: The X and Y chromosomes in human beings that determine the sex of an individual. Females have two X chromosomes in diploid cells; males have an X and a Y chromosome. The sex chromosomes comprise the 23rd chromosome pair in a karyotype. Compare autosome.

Shotgun method: Cloning of DNA fragments randomly generated from a genome. See library, genomic library.

Single-gene disorder: Hereditary disorder caused by a mutant allele of a single gene (e.g., Duchenne muscular dystrophy, retinoblastoma, sickle cell disease). Compare polygenic disorders.

Somatic cells: Any cell in the body except gametes and their precursors.

Southern blotting: Transfer by absorption of DNA fragments separated in electrophoretic gels to membrane filters for detection of specific base sequences by radiolabeled complementary probes.

Stochastic effect: A random effect which may occur when sampling low levels of DNA extract for the PCR reaction. When this occurs not all the different alleles in the extract are sampled to the same extent. The result is that a particular allele is not amplified as well as the other alleles, or is not amplified at all.

STR (short tandem repeat): A locus which is composed of a short sequence of DNA which is repeated a number of times.

STS: See sequence tagged site.

Subpopulation: A subgroup of people who are not mating randomly with the broader population.

Substrate: any surface on which a layer of different material can be deposited.

Supernatant: the liquid floating above or on the surface of matter deposited by centrifugation, precipitation or other means.

Tandem repeat sequences: Multiple copies of the same base sequence on a chromosome; used as a marker in physical mapping.

Technology transfer: The process of converting scientific findings from research laboratories into useful products by the commercial sector.

Telomere: The ends of chromosomes. These specialized structures are involved in the replication and stability of linear DNA molecules. See DNA replication.

Thymine (T): A nitrogenous base, one member of the base pair A-T (adenine-thymine).

Transcription: The synthesis of an RNA copy from a sequence of DNA (a gene); the first step in gene expression. Compare translation.

Transfer RNA (tRNA): A class of RNA having structures with triplet nucleotide sequences that are complementary to the triplet nucleotide coding sequences of mRNA. The role of tRNAs in protein synthesis is to bond with amino acids and transfer them to the ribosomes, where proteins are assembled according to the genetic code carried by mRNA.

Transformation: A process by which the genetic material carried by an individual cell is altered by incorporation of exogenous DNA into its genome.

Translation: The process in which the genetic code carried by mRNA directs the synthesis of proteins from amino acids. Compare transcription.

tRNA: See transfer RNA.
U

**Uracil:** A nitrogenous base normally found in RNA but not DNA; uracil is capable of forming a *base pair* with *adenine*.

V

**Vector:** See *cloning vector*.

**Virus:** A noncellular biological entity that can reproduce only within a host cell. Viruses consist of *nucleic acid* covered by *protein*; some animal viruses are also surrounded by membrane. Inside the infected cell, the virus uses the synthetic capability of the host to produce progeny virus.

**VLSI:** Very large-scale integration allowing over 100,000 transistors on a chip.

Y

**YAC:** See *yeast artificial chromosome*.

**Yeast artificial chromosome (YAC):** A vector used to clone DNA fragments (up to 400 kb); it is constructed from the telomeric, centromeric, and replication origin sequences needed for replication in yeast cells. Compare *cloning vector, cosmid*. 