Produced by the Centre for Genetics Education. Internet: http://www.genetics.edu.au

Important points

- The chance that two unrelated people have an identical total DNA gene sequence is at least 1 in 6 million
- The closer two people are related to each other, the greater the chance will be that their DNA sequence will be similar. Identical twins have almost exactly the same DNA sequence
- The small differences and similarities in the DNA sequence between people are used in genetic testing to determine the identification of individuals
- The tests do not look at the information in the genes but instead examine non-coding DNA that separates the genes along the chromosomes
- Specifically, the DNA is examined where there are large numbers of **repeated** sequences of letters along its length eg. ATTCGATTCGATTCG
- As each person has two copies of each chromosome, they will each have two copies of a specific number of repeats of sequences (usually three to five letters in length; called *short tandem repeats - STRs*)
- The pattern of different numbers of STRs at certain sites on the chromosomes are used to create a DNA pattern or **DNA fingerprint** that is as unique as possible for each person
- The number of repeats in each pattern for each person is measured and is used to create a numerical **DNA profile** for that person
- Identification DNA testing using DNA profiling is used as an aid in identifying victims of crime, natural or other disasters such as the terrorist attacks, suspects in a crime and in determining paternity and kinship for immigration purposes for example
- Increasingly, DNA obtained from crime scenes is first checked to see if it matches with DNA profiles stored in databases. The higher the probability that it matches, the greater the chance that they are from the same person
- Limitations and concerns with DNA profiling and databanks of profiles include:
  - DNA fingerprinting still gives only a chance or probability that two samples are from the same person
  - The generation of the profile also depends on how common the different patterns of repeat numbers at the loci in different population groups are
  - Privacy and ethical concerns with the DNA databanks and their possible misuse
  - The concerns with widespread promotion and applications of biological relationship testing including the necessity for informed consent by both parents for a sample to be taken from a child, quality control, the availability of counselling after the test result and privacy, particularly in relation to mail order paternity tests

Genetic testing described in Genetics Fact Sheet 21 involves the analysis of the information in the 'coding DNA' that makes up genes located on specific chromosomes.

If a chromosome is considered to be like a string of beads where the beads are the genes, the DNA 'string' between the genes is called 'non-coding' DNA.

- 'Non-coding' DNA does not contain the coded messages that the cells use and sometimes has been erroneously referred to as 'junk' DNA
- It is increasingly clear that the role of this form of DNA is important, perhaps in controlling how the genes work
- 75% of the total DNA in the human genome is 'non-coding' DNA between and within the genes

Analysis of the non-coding DNA between the genes has applications in forensic studies and biological relationship testing.

- The genetic test utilises the many small differences that are present in this DNA (called *polymorphisms*) between individuals to create a fingerprint for each person based on their genetic code

The DNA sequence - differences and similarities between humans

The genetic code in all DNA in humans is made up of a string of 6 billion or so 'letters': A, T, C and G (see Genetics Fact Sheets 1 &4). These letters are chemical 'bases' of the DNA molecule.

In the 'coding DNA' which makes up most of the DNA in the genes, the letters are combined into groups of three to produce a message made up of a sequence of three letter words. The genetic information instructs the cells to enable the body to grow, develop and function.

The non-coding DNA is of course also made up of a long string of the letters A, T, C and G.

- An important feature of this non-coding DNA is that it contains large numbers of **repeated** sequences of letters along its length eg. ATTCGATTCGATTCG
- The number of times that a sequence is repeated within a length of non-coding DNA can range from just a few to hundreds
- Studies of these repeated sequences are proving very useful in tracking our evolutionary past (see Genetics Fact Sheet 24)

As we have evolved, changes have built up in our non-coding DNA as well as our genes. While changes in the coding DNA sequences in the genes may or may not cause a problem with how the genes work, many changes to the non-coding DNA have occurred with no impact on the individual.

Despite these changes that have built up in the non-coding DNA

- Most humans are overwhelmingly alike in their sequence of letters in their coding DNA ie. in their genes, regardless of race or ethnicity although each person's sequence is unique (except for identical twins)
- DNA samples from two unrelated people differ, on average, at only one letter or 'base' per thousand
- That works out to be 1/1,000th of 6 billion that equals 6 million: ie. in their total DNA, two unrelated people have 6 million differences in their sequence of letters
- This is enough to produce all the genetic differences between these two people
- The chance that two unrelated people have an identical total DNA gene sequence is at least 1 in 6 million
- The closer two people are related to each other, the greater the chance that their DNA sequence will be similar
- Identical twins have exactly the same DNA sequence

The small differences and similarities in the DNA sequence between people are used in the tests to determine the identification of individuals as described below.

Testing the non-coding DNA for non-medical purposes

The testing is done using the system described in Genetics Fact Sheet 21 to look at particular genes. Unlike tests on DNA for medical purposes however, here the tests look at specific areas (called *loci*-singular *locus*) in the non-coding DNA that have nothing to do with neither how our bodies grow and develop nor our health.

These loci are found at a number of sites on **each** chromosome. The loci that are most useful are those that have been found to have different numbers of the repeated sequences in different people in the population. These sites are described as being 'polymorphic' that means 'many forms'.

Each person has two copies of each chromosome. So each person will have a specific number of repeats at each of these loci on one of their chromosomes and the same, or different, number of repeats, at the partner locus on the other chromosome.

These different forms at the same sites on the partner chromosomes are called *alleles* of the locus. This information is used to create a DNA pattern or DNA fingerprint that is as unique as possible for each person as described below.

Creating a DNA fingerprint for a person

Step1:

- In the laboratory, using enzymes that are chemical 'scissors', the DNA is cut into hundreds of small pieces (*Figure 22.1*) at sites where there are specific sequences of the DNA letters (usually four to six letters in length)

- As everyone's DNA has some small differences, the sites may be at different places in people's non-coding DNA and so the enzymes will cut the DNA into different sizes in different people

Step 2:

- The cut DNA is placed into a slab of 'jelly' (a gel matrix) and an electrical current is applied so that the 'jelly' becomes electrically charged and has a 'positive' (+) end at the top and a negative (-) end at the bottom, just like the positive and negative ends of a battery
- As the DNA is a chemical which has a negative charge, the DNA is pulled towards the positive end of the gel or from the top to the bottom
- The pieces of DNA separate along the gel according to size: the biggest pieces move the slowest and so will remain near to the top of the gel
- The gel now contains all of the person's DNA spread from the top to the bottom of the gel

Step 3:

- To select out the pieces of DNA that need to be looked at, the pieces of DNA that have spread through the gel are covered with special DNA 'probes'
- The probes are made in the laboratory and contain a match for the DNA sequence that the test is looking for
- The probes in fact have the opposite sequence to the sequence being tested for. They match up because of the ability of the letters A and T, and C and G to pair with each other as shown in *Figure 22.2*
- If the person's DNA on the gel contains the matching sequence, the probe will combine (*hybridise*) with the person's DNA at the site of the matching sequences
- Chemicals such as fluorescent dyes are attached to the probes. When the gel is exposed to a certain type of light, the DNA on the gel that has been detected as containing a particular sequence of 'letters' will show up. The remaining DNA will not be visible

When a number of different probes are used to detect different repeats of sequences, the gel will look like a series of bands as
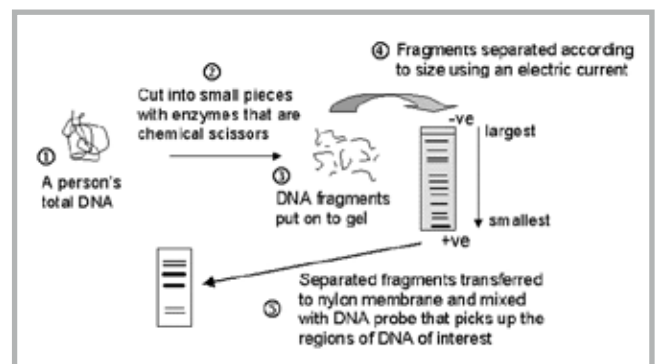

Figure 22.1: *Genetic testing to create a DNA fingerprint of a person.*

shown in *Figure 22.3*. Each band represents a site or locus on the non-coding DNA at which a specific repeated sequence is located.

- Some of the bands appear darker than others. This is because that person has the same number of repeated sequences of letters on each of their two chromosomes where the locus is located. The band is showing the same pattern from each chromosome and so the two bands are lying on top of each other on the gel

- A person can only ever have one or two bands for each probe

- Persons 1 and 2, however, have the same banding pattern for probe 5 and persons 2 and 3 have the same pattern for probe 7

- That is why it is necessary to use a number of different probes to enable differentiation between different people

- The bands at the bottom of the gel represent shorter sequences of letters than at the top. For example, a band at the top of the gel may represent a sequence of five letters repeated 30 times so it will be 150 letters or bases long. At the bottom the sequence may be made up of only four letters repeated 10 times so it will be only 40 letters or bases long

These patterns of bands have become known as a person's **DNA fingerprint**, an analogy with the classical fingerprint system of identification.

DNA fingerprinting in forensic investigations

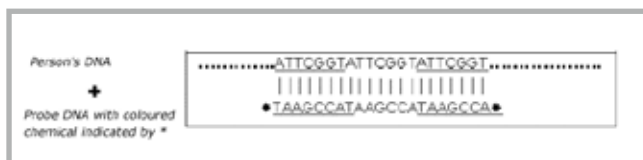Genetic markers have long been used to identify people, even prior to the use of DNA fingerprinting.

Figure 22.2: *Using a probe to detect a repeated sequence in the non-coding DNA of a person.*
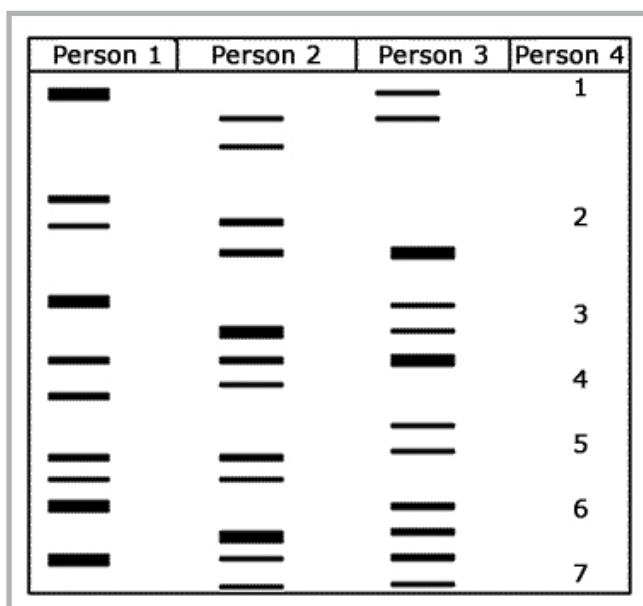

Figure 22.3: *DNA fingerprints of three people created using seven different DNA probes manufactured in the laboratory.*

The first genetic markers used were the ABO blood groups (A, B, AB and O) but over the years, other blood grouping systems were discovered, such as the MN system, and by the 1960s there were 17 different blood group systems known.

Not all were suitable for forensic studies but by the 1980s a variety of these blood groups and a number of proteins and enzymes were used in a battery of forensic tests.

In 1985, however, the analysis of DNA to produce DNA fingerprints was introduced to increase the accuracy of the identification. By using a number of different regions of the DNA (loci) where the tandem repeated sequences occurred (5-6), the chance or probability that two samples were the same was calculated. A match probability was calculated.

In that year, the analysis of segments of repeated sequences at these different loci in the non-coding DNA was used in a criminal case of rape in England. The DNA obtained from the rape victims, the semen and from the suspects was cut into small pieces and put onto a gel as described above with manufactured DNA probes of these sequences.

As a result, one suspect in two rape homicides of young girls was cleared and the rapist was found, on the basis of a DNA fingerprint similar to that shown in *Figure 22.3*.

This DNA fingerprinting system was considered preferable to the earlier systems. The DNA samples are very stable (DNA has even been examined from Egyptian mummies), can come from a variety of sources as the DNA is the same no matter what cell type and the individual variability is much greater than revealed by the blood groups or other protein samples.

From DNA fingerprinting to DNA profiling in forensic studies

While DNA fingerprinting enabled matching of samples with a high probability from crime scenes with those of suspects in many cases, there were often difficulties in visually separating bands that were very close together.

The loci used were called **variable number of tandem repeats** (VNTRs), usually 8-80 letters in length so that separation was sometimes not optimal.

The analysis of VNTRs is now being replaced by the analysis of repeats of shorter sequences of letters called **short tandem repeats** (STRs) that are usually three to five letters in length.

- STRs can be used even when the sample of DNA is very small such as that from a cigarette, hairs, coffee cups or the saliva from a postage stamp

- Their interpretation is also usually less uncertain and the analysis takes only a few days instead of weeks using VNTRs

- Automatic systems are being developed to make the analysis even more rapid

The more regions of the DNA where the STRs are located (loci) are used, the greater the chance that variability will be found between unrelated people. In addition, analysis of markers on the two sex chromosomes (X and Y) immediately identifies the sex of the person. Further information can be gained from the analysis of the DNA in the mitochondria, small units in the cells that contain genes involved in energy production (see Genetics Fact Sheets 1 and 12).

DNA GENETIC TESTING - PATERNITY AND FORENSIC USE

Table 22.1. Profiler Plus is used in Australian forensic laboratories and is made up of nine different markers in the non-coding DNA on various chromosomes as well as a marker on the X and Y chromosomes, to give a match probability.

STR marker in Profiler Plus	D3	VWA	FGA	D8	D21	D18	D5	D13	D7	AMEL
Repeat numbers of sequences of letters observed in the Caucasian population	12, 13,14, 15, 16, 17,18, 19	12, 14, 15, 16, 17, 18, 19, 20	18, 19, 20, 21, 22, 23, 24, 25, 26, 27	8, 9, 10, 11, 12, 13, 14, 15, 16, 17	26, 27, 28, 29, 29.2, 29.3, 30, 30.2, 31, 31.2, 32, 32.2, 33.2, 34.2	10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22	9, 10, 11, 12, 13, 14, 15	8, 9, 10, 11, 12, 13, 14, 15	7, 8, 9, 10, 11, 12, 13, 14, 15	X, Y

- Each STR locus has a number of different possible patterns according to the different number of repeats of the sequence, but each person will only have two of all the possible patterns in their DNA
- Rather than looking at the bands generated from these patterns, and creating a visual fingerprint, the number of repeats in each pattern for each person is measured
- This gives a numerical **DNA profile** for these STRs for each person as described below
- Currently, Australian forensic laboratories use a bank of STRs called *Profiler Plus* that is made up of nine different loci in the non-coding DNA on various chromosomes as well as a locus on the X and Y chromosomes to give a match probability. In addition, examination of mitochondrial DNA is carried out as necessary

Each of the loci in Profiler Plus has a number of different repeats of each sequence that have been observed in the population (*Table 22.1*). As each person can only have 1 or 2 of these forms (sometimes the locus on both chromosomes have the same number of repeats), there are large numbers of possible combinations, increasing the capability of the system to differentiate between individuals in the population.

For example, the *D3* STR locus is a DNA segment that contains the letters AGAT repeated between 12 and 19 times in different people:

- Some people will have 12 repeats of AGAT on one chromosome and have 16 repeats of this sequence on the paired chromosome
- Other people may have 14 repeats of the letters AGAT in both copies of this STR locus

Table 22.2 shows the DNA profiles for 2 people. Person 1 is a female and person 2 is a male, identified by the *AMEL* STR locus on the sex chromosomes. Using this system, comparisons can be made between individuals and samples of DNA obtained from a crime scene to give a probability of a match.

DNA in the courtroom

DNA profiling is used as an aid in identifying victims of crime, natural or other disasters such as the terrorist attack on the World Trade Centre on September 11 2001 and suspects in a crime. In some cases the test can exclude a person as a suspect and reverse a previous criminal conviction.

For example, in Australia a conviction for murder and rape was made 28 years after the crime had been committed: DNA obtained in 1983 from semen stains found on a towel that had covered the victim's body was matched with the DNA of another man. The jury was told that the probability that another person could have had the same DNA profile was 1 in 43 trillion. In the United States, over 100 people have been released from 'Death Row' after acquittal based on DNA evidence.

Table 22.3 is an example of the use of DNA profiling to identify an offender in a rape case from a number of suspects.

Figure 22.4 shows how the results are generated automatically for three of the STR markers: *D3, VWA* and *FGA*.

Increasingly DNA obtained from crime scenes is first checked to see if it matches with DNA profiles stored in databases.

- These databases contain the DNA profiles obtained from convicted criminals, from other crime scenes and suspects
- A link between an offender or suspect and a crime scene is known as a 'cold hit' because it allows the police to `identify' a suspect by his or her DNA profile alone where there are no other leads available in the case
- Therefore, the database could potentially allow police to solve a number of unsolved crimes

Forensic DNA databases have been established in the United Kingdom, the United States, Canada and New Zealand and in
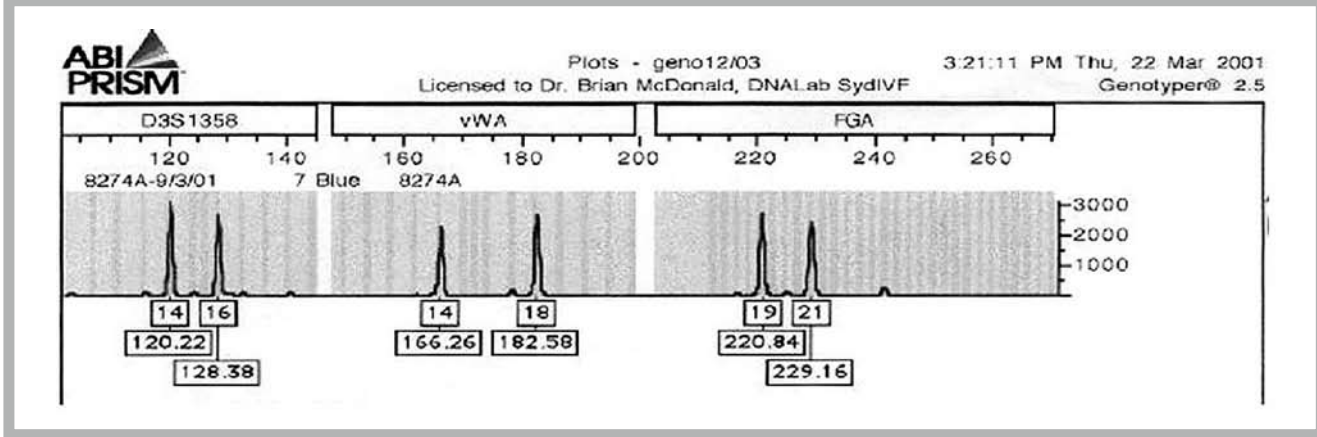
Table 22.2. DNA profiles for two people. Person 1 is a female and person 2 is a male, identified by the AMEL STR locus on the sex chromosomes.

STR marker in Profiler Plus	D3	VWA	FGA	D8	D21	D18	D5	D13	D7	AMEL
Person 1	14, 14	15, 18	19, 20	11, 13	31.2, 34.2	16, 19	11, 11	10, 12	10, 10	X, X
Person 2	17, 18	17, 17	18, 22	13, 15	28, 30	10, 14	11, 11	9, 14	11, 13	X, Y

Table 22.3. Identification of suspect 2 as an offender in a rape case.

STR marker in Profiler Plus	D3	VWA	FGA	D8	D21	D18	D5	D13	D7	AMEL
Victim	14, 15	16, 18	19, 20	11, 13	30, 30	16, 19	11, 12	9, 12	10, 10	X, X
Semen specimen	16, 18	17, 17	18, 22	13, 15	28, 30	10, 14	11, 11	9, 14	11, 13	X, Y
Suspect 1	14, 15	16, 18	19, 20	11, 13	28, 30	19, 19	11, 12	12, 12	10, 11	X, Y
Suspect 2	16, 18	17, 17	18, 22	13, 15	28, 30	10, 14	11, 11	9, 14	11, 13	X, Y
Suspect 3	14, 15	16, 18	18, 20	11, 15	28, 29	14, 16	12, 12	12, 14	10, 13	X, Y

Figure 22.4: Computerised generation of a DNA profile (Courtesy of Dr Brian MacDonald, DNA Labs, SIVF).

Australia by the Crimes Amendment (*Forensic Procedures) Act 2001 (Cth) (Forensic Procedures Act)*, which came into force on 20 June 2001.

The *Forensic Procedures Act* inserted more detailed provisions into the Crimes Act in relation to the carrying out of forensic procedures, and expanded the scope of the coverage regarding volunteers and serious offenders.

It also provided a legislative framework for the operation of a national DNA database system known as the National Criminal Investigation DNA Database (NCIDD). The database is operated by CrimTrac, an executive agency established under the *Public Service Act 1999* (Cth).

Limitations and concerns with DNA profiling and databanks of profiles

DNA fingerprinting still only gives a chance or probability that two samples are from the same person.

In forensic cases, the police usually have DNA from the crime scene and are seeking the chance that there is a match with the DNA from a suspect. They are given a 'match probability' or the chance that the two samples match. The higher the probability, the greater the chance that the samples of DNA belong to the same person.

- The generation of the profile also depends on how common the different patterns of repeat numbers at the loci are in population groups. For example, the range of repeats at the STR loci in those who are of European ancestry may be different to the range seen in people of Asian origin
- There are a number of concerns about the use of DNA forensic material in the courtroom and the establishment of DNA databanks

These include consent for the sample, methods of its collection and procedures to ensure that it is not contaminated from the time it is taken to when it is tested. Other concerns relate to storage of the sample, its destruction after testing and the privacy and confidentiality of the profile generated.

An Australian Federal Government inquiry conducted jointly by the Australian Law Reform Commission (ALRC) and the Australian Health Ethics Committee (AHEC) into the protection of Human Genetic Information in Australia (http://www.alrc.gov.au) has examined these issues and has made extensive recommendations in its final report 'Essentially Yours' (2003).

DNA profiling in determining validity of biological relationships

Increasingly DNA testing is being undertaken to determine parentage or other family relationships.

Until recently, such testing was primarily paternity testing: that is, establishing whether a man had fathered a child. *Table 22.4* is an example of the results of a DNA parentage test.

Just like in forensic testing which originally relied upon blood groups and then moved into DNA fingerprinting, DNA profiling is increasingly used in kinship testing as well.

- A child must be a combination of the DNA of its mother and father
- Using DNA profiles of all the people involved, usually mother, child and two men questioning their fatherhood, a probability can be given for one of the men being the father
- A man can be excluded as the father if he does not match with the child on at least two STR loci. Inclusion as a father is preferably with 99.9% probability

Table 22.4: Example of DNA profiling used in paternity testing. Possible Father 1 would be excluded but there is a high probability that Father 2 is the true father.

STR marker in Profiler Plus	D3	VWA	FGA	D8	D21	D18	D5	D13	D7	AMEL
Mother	14, 16	16, 18	19, 20	11, 13	30, 30	16, 19	11, 12	9, 12	10, 10	X, X
Child	14, 14	16, 18	19, 20	11, 13	30, 30	16, 19	11, 12	9, 12	10, 10	X, X
? Father 1	16, 18	17, 17	20, 22	13, 15	28, 30	14, 16	11, 11	9, 14	11, 13	X, Y
? Father 2	14, 15	16, 18	18, 20	11, 15	28, 29	14, 19	12, 12	12, 14	10, 13	X, Y

- For parentage analysis, a paternity index (PI) is produced. This is a measure of the probability that compares the chance that the man is the father in the mother-child-father combination to the chance that the man was randomly chosen from the population

- The PI will either exclude a man as the father or demonstrate that there is a high probability that he is the father of the child

Concerns with widespread promotion and applications of biological relationship testing

Paternity is often challenged in court proceedings when child maintenance and custody, and often succession to property, is at issue. The results of a 'parentage testing procedure' will only be able to be considered by the Family Court if undertaken by a laboratory accredited for paternity testing. Paternity tests, however, are now available commercially, by mail order and through the internet.

This development raises issues that again were of concern to the ALRC Federal Inquiry. These included the necessity for informed consent by both parents for a sample to be taken from a child, quality control, the availability of counselling after the test result and privacy, particularly in relation to mail order paternity tests.

The ease with which a sample from which DNA can be extracted can be obtained from a child by one parent without the other's knowledge, and the marketing of commercial tests make these issues worthy of urgent consideration.

Other Genetics Fact Sheets referred to in this Fact Sheet: 1, 12, 21, 24