

Criminal and Family Law Issues re:
DNA Evidence

By: Andrew Mason, B.A. LL.B. of the Saskatchewan Bar

Prepared for Saskatchewan Legal Education Society Inc. (SKLESI)
April 17, 2009

1.0 Introduction

As we are reminded with each CSI episode, DNA evidence has revolutionized forensic investigation and proof.

DNA is most often used to establish identity in criminal cases and to establish paternity in family cases. In identity cases, DNA evidence is said to be “conclusive”. DNA experts throw out astronomical probabilities, like one in 100 billion, as the odds against a false match. In establishing paternity, the odds are not as high, but still persuasive.

It is important that counsel understand the fundamentals of DNA technology and become able, with the assistance of a qualified expert, to verify the accuracy of the results or identify possible weaknesses in the interpretation of the results.

In this paper I will attempt to explain the essential science behind DNA “fingerprinting”, how reliable it really is as proof of identity or paternity, and provide the practitioner with some guidance in how to deal with the evidence and the conclusions that flow from it.

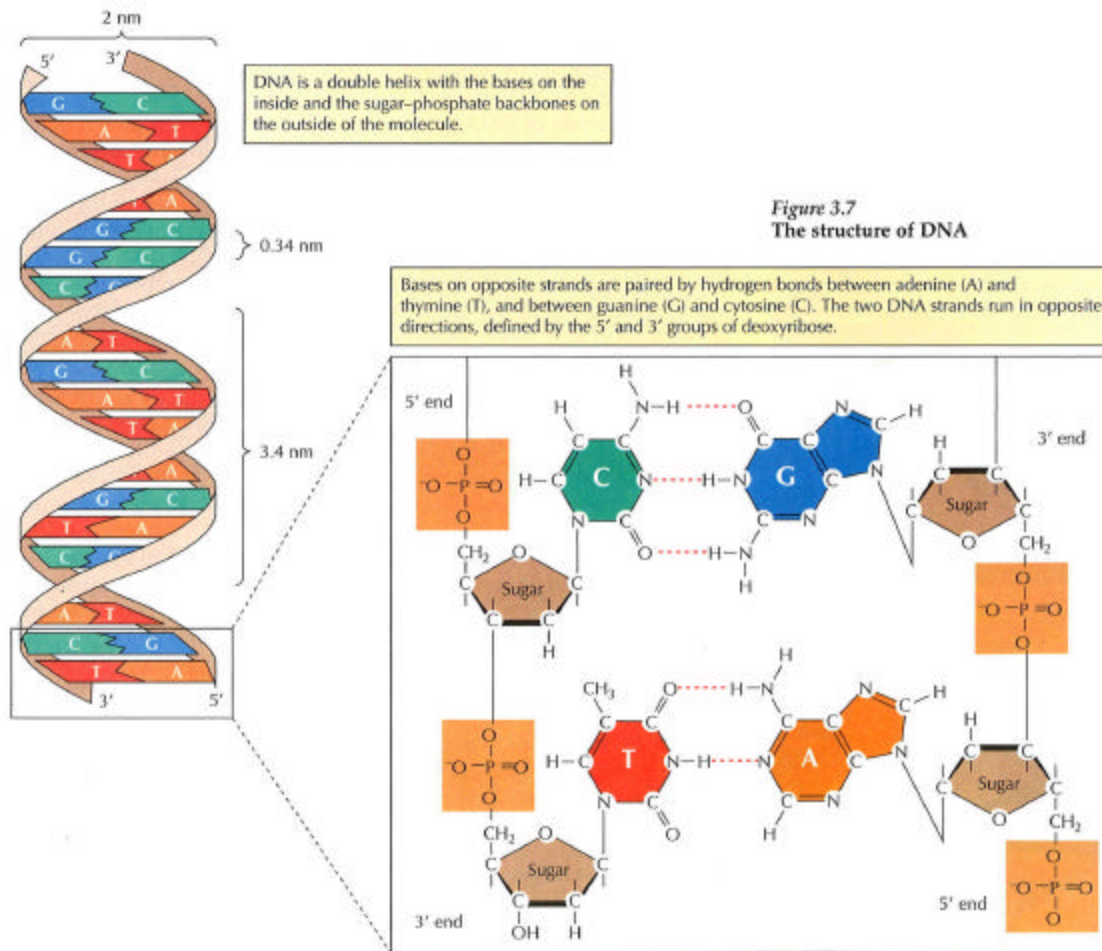
2.0 DNA fingerprinting

2.0.1 The science of DNA

A DNA “fingerprint” is a profile of 14 areas of the DNA of a cell. All cells of the human body have 23 pairs of chromosomes in the nucleus (except for sperm/egg cells, which have 23 single chromosomes each, and red blood cells, which have no nucleus at all.). Each cell also contains mitochondria organelles situated in the cytoplasm of the cell outside the nucleus which contain small amounts of DNA. The mitochondrial DNA comes from the mother’s egg cell only so it is highly conserved within populations. It does not provide a unique code for each person.

The cell is really an elaborate manufacturing complex that produces protein molecules whose properties depend upon the sequence of amino acids of which they are comprised. The machinery involved in the production of these proteins uses information contained in the nuclear DNA as “blueprints” for manufacturing these protein molecules. DNA can be thought of as a “code” made up of the letters A, C, G and T¹. This DNA code directly determines the amino acid sequence of the protein molecules produced by the cell.

¹ These letters represent the four molecules, called nucleotides or nucleotide bases, which join together with a molecular backbone made of sugar/phosphates to form the DNA polymer molecule. A



The structure of DNA (from *The Cell, A Molecular Approach* by G. M. Cooper, (2nd ed.) ASM Press, 2000 at p. 95)

The “DNA fingerprint” concept is based on the recognition that although 99% of the human genome (the 23 chromosome pairs) does not vary significantly between individuals, there are certain regions of the nuclear DNA which do. In particular, there are regions of the DNA containing particular sequences which are repeated a number of times in a row. These regions are known as VNTRs (Variable Number of Tandem Repeats) or STRs (Short Tandem Repeats).

Years of scientific study have shown that there are many such VNTR or STR regions, referred to as *markers* or *loci* (plural of locus) that are highly variable between individuals². That is to say the length of such regions (the number of repeats of the

stands for Adenine, C for Cytosine, G for Guanine and T for Thymine, which are the four nucleotide bases in DNA. The entire genome of a cell contains approximately 3 million such “letters” (and, if stretched out in a line, would be 3 metres long).

² See *infra*, section 3.0 on DNA databanks.

known sequence) may have many (usually in a range from 5 to 30) different possibilities. Moreover, the variability of length of one marker is completely independent of the variability of length of the other markers.

It is the independent variability of these particular well-studied regions that gives each human being a particular “signature” or “fingerprint”. If one looks at only one or two such loci, there will be many people who share a person’s DNA profile. But if one examines many such loci, the chance of two people having exactly the same DNA profile becomes vanishingly small³.

2.0.2 Obtaining the DNA fingerprint

The DNA “fingerprint” is a remarkable scientific tool that is built on decades of scientific achievement.

The early DNA profiling technique, known as the RFLP⁴ method, required a DNA sample that was large enough to provide a result that was visible to the eye. Literally millions of cells were required in a sample to provide a DNA fingerprint. Since a small quantity of human fluids may contain several million cells, this was not always a problem. But in many cases where the only crime scene residue consisted of a few skin cells, it was a problem.

Since the mid to late 1990’s DNA profiling has been done using a technique known as PCR⁵ which first makes millions of copies of the DNA regions of interest. With PCR technology, a DNA fingerprint can be obtained from the genome of a single human cell.

The PCR method uses DNA sections called primers that identify and adhere to the loci of interest. Specialized DNA replicating molecules are added and the mixture is cycled through several different temperatures for periods of time. This results in copies of the DNA of only the loci of interest.

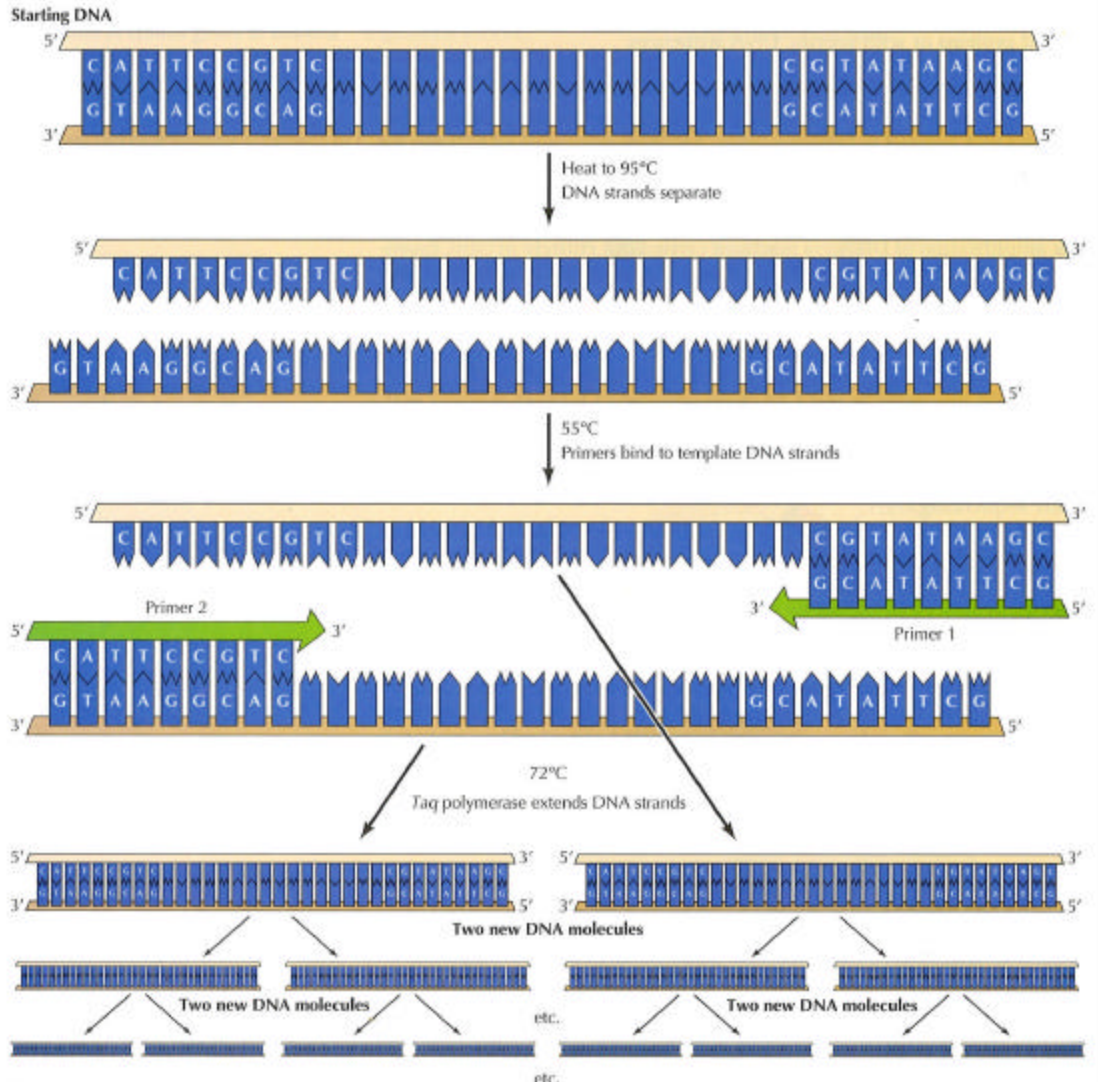
PCR “amplification” is explained in many resources on the internet. A good video on PCR may be viewed at: <http://www.youtube.com/watch?v=j9Gu7iwBi4I>

The following diagram illustrates how PCR works.

³ This is just simple common sense but it can also be shown with mathematics. If we identified a person by hair colour and birth month only, many would share the same profile. If there were 10 possible hair colours, the probability that another person would have the same profile would be $1/10 \times 1/12 = 1/120$. But if we identified a person by hair colour, birth month, birth day, birth year, eye colour, height, sex, first initial, last initial, middle initial, and community name, the number of people sharing that same profile becomes very small. Assuming that these characteristics are independent and quite random, the probability that another person would have the same profile would be the product of all the probabilities of a match on each one – which we can see would be an extremely small number.

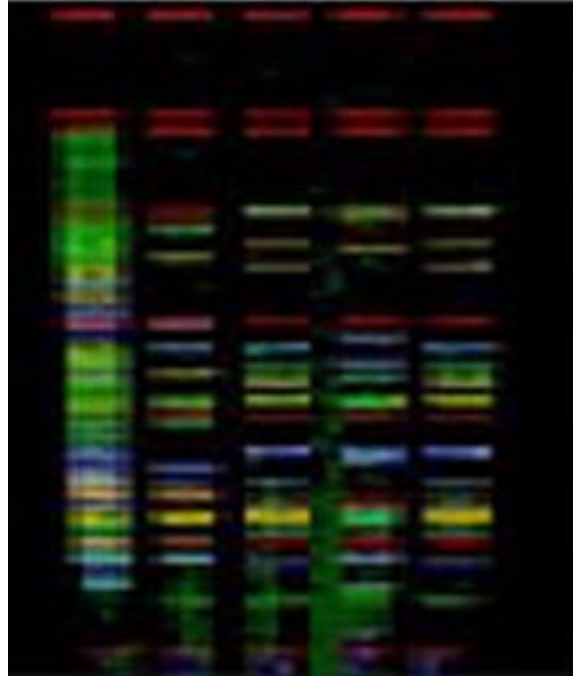
⁴ RFLP stands for Restriction Fragment Length Polymorphism.

⁵ PCR stands for Polymerase Chain Reaction.



PCR amplification (from *The Cell, A Molecular Approach* by G. M. Cooper, (2nd ed.) ASM Press, 2000 at p. 116)

Once the DNA of interest (a standard 14 specific markers are used) is amplified using PCR technology, the next step is to separate out the DNA molecules according to length (using a process called electrophoresis). Primers embeded with fluorescent dyes are often used (with different colours for different markers) in the PCR process to enable us to better see and interpret the results. The shortest molecules travel the farthest and the longest travel the least (from top to bottom), as seen in the following photograph:



A DNA fingerprint of 4 individuals (right four columns).
(from the National DNA Databank web site at:
http://www.nddb-bndg.org/techno/graph_e.htm)

2.0.3 Interpreting the Results

The length of tandem repeats in the regions of DNA being examined are not unique. They are actually quite common – some regions may have only 5 or 10 length variations so the chance of another individual chosen at random from the population having the same “bar” match for a particular marker may be as high as 20%. The key however, is the evidence that the lengths of these regions of DNA vary widely within the population AND that the length of DNA of one marker is independent of the length of another.⁶

The DNA profile results from taking many markers (usually 14) and comparing their relative lengths. The probability that two people will have the exact same lengths of DNA sequences at each locus is the product of 14 probabilities, each of which is about 1/5 to 1/10. That results in a very small number: in the order of $1:5^{14}$ or 1 in 30 billion to $1:10^{14}$ or 1 in 100 trillion

3.0 Standardization of DNA Profiles: CODIS and DNA Databanks

In order to compare DNA collected at the scene of a crime with the DNA profiles obtained previously from a large population, it is essential that there be a standard DNA profile and a system for measuring and comparing profiles.

⁶ This is true only in the case of unrelated individuals. DNA markers are not randomly distributed within populations of related individuals.

The FBI in the U.S. began standardizing DNA profiles in 1994 with the Combined DNA Indexing System (CODIS). CODIS employs 13 markers, plus one marker to determine sex⁷. Virtually all criminal DNA testing for criminal identification purposes now conforms to the CODIS standard.

The U.S. and the U.K. maintain the largest DNA databanks in the world, each containing millions of DNA profiles. Canada's National DNA databank began in 1998 with the passage by Parliament of the **DNA Identification Act**.

4.0 DNA fingerprinting and Parental Testing

When using DNA profiling to establish the identity of a parent, the same DNA fingerprinting technique is used. There is no requirement to adhere to the CODIS standard set of markers, though these are often used. The choice of markers may depend on the circumstances. For example when testing the paternity of a male child, certain loci on the y-chromosome may be examined since the y chromosome comes entirely from the father.

The number of tandem repeats of a sequence at a given locus is determined by heredity. For a given locus a person carries two alleles or variations, one from the mother and one from the father. Each parent also carries two alleles, so there are $2 \times 2 = 4$ different possible combinations of alleles for each locus in the child's DNA. (There is also the small possibility that a child will not carry an allele from either parent because of the chance (about 1:200) that a random mutation can occur in the DNA from one generation to another).

Absent genetic mutation between parent and offspring, a child will have, for each locus, two alleles, (sections of DNA of the tandem repeat sequence) each with a tandem repeat length that matches a section from one of each of the child's parents). Since for any single locus there are four possible ways of combining the parent's alleles in the child, the chance of two offspring of the same parents having the same alleles at a particular locus is $\frac{1}{4}$. So the chance that two children of the same parents (not identical twins) will have identical DNA profiles (all 14 loci the same) is very small.

Paternity is determined by examining the results for each allele from the putative parents and the child. If, for each locus, the 2 alleles from the child match with one allele from the mother and one from the father, there is a very high probability that these are the correct parents.

⁷ The technical names for the CODIS markers are D3S1358, THO1, D21s11, D18s51, D5s818, D13s317, D7s820, D16s539, CSF1PO, vWA, D8S1179, TPOX, FGA. The marker for sex is AMEL.

For an example of a DNA fingerprint for paternity testing, see Appendix 1. Note how for each locus there are two bands, one identical to one from the mother and the other identical to one from the father.

The Reliability of DNA Fingerprints

5.0.1 In Criminal Identification cases

A DNA “match” or “no match” between the profile taken from a crime scene sample and that of a suspect is highly probative. Exactly what it is probative of depends on surrounding facts. The reliability of any conclusion that may be drawn from the “match” or “no match” depends on those facts.

For example, if the DNA profile from semen obtained from a murder victim matches the accused’s DNA profile, is highly probative that the accused had sexual relations with the victim. Whether that can be used to reliably conclude that the victim was the murderer depends on other facts, such as the kind of relationship between the accused and the victim, whether there is evidence that the murder occurred during a sexual assault and whether there is DNA from another person found on the victim or at the scene of the crime.

It is highly recommended that a qualified DNA expert be retained by counsel to review the DNA evidence.

Sample errors: It is very important to verify that there was no error in DNA sampling. Was the sample labeled as your client’s really provided by your client? Given the rigorous protocols used by police, this is not likely to occur. But it is a relatively easy thing to verify. The DNA “match” can be checked by having a second DNA sample taken from the client and tested in an independent laboratory using the same DNA probes used by the police laboratory.

Check for unusual features: A DNA expert may review all of the DNA results from the victim, the accused and other suspects. It may appear from some of the results that a third person’s DNA was present. Samples that may have been rejected on the grounds that some of the DNA has been degraded, may in fact have some significance.

Check for evidence of Contamination: Contamination of samples with small amounts of human DNA is a potential problem with PCR since the DNA from a single cell will result in millions of copies being made and then showing up in the resulting profile. Such contamination is usually evident from the results themselves, but there may be instances where what may be thought to be contamination is significant (eg. showing the presence of a third person at the scene). The expert will be able to provide assistance in interpreting such results.

At the end of the day, however, if the DNA samples were correctly identified and show no unusual indicators, your DNA expert will tell you that the results are very reliable.

This is not a matter of debate or dispute within the scientific community. The science is valid.

Much the same way that a photography expert will tell you that a film clearly showing your client's face at the scene of the crime means that the film was developed properly and that your client was there, DNA evidence often speaks for itself.

5.0.2 The Reliability of DNA Fingerprints in establishing Parents

As explained above (section 4.0), using DNA profiles to verify parental relationships does not involve "matching" DNA. It involves comparing DNA profiles of the putative genetic relatives of the subject to the DNA profile of the subject.

The conclusion as to parenthood from DNA profile comparison has a low margin of error but is not quite as "certain" as DNA matching. It can be used to determine whether an individual is a parent of a child to an accuracy of 99.99% (ie. one error in 10,000). Because there is a small possibility of random mutation in the number of tandem repeats, it is possible that one or more of the "bars" may not match up with a real parent. For this reason, exclusion of a parent should be based on several mismatched markers.

Again, it is highly recommended that a qualified DNA expert be retained by counsel to review the DNA evidence.

6.0 Statistical fallacies

6.0.2 The Arizona Search

In 2001, a laboratory scientist was examining the profiles in Arizona's DNA databank and discovered that in the database of 65,000 persons were several instances in which persons had a 9 marker DNA "match". This gave rise to great concern since it seemed to contradict the 1 in 30 billion numbers that were being touted. See Appendix 2.

It must be noted that in a DNA databank, which may contain many related individuals, the probability of finding two entries that match on 9 of 14 markers is a much larger probability. This does not diminish the probability of a match with respect to a particular profile of 14 markers in a population of unrelated individuals.⁸

For a general discussion of the probabilities involved, see:

<http://dna-view.com/ArizonaMatch.htm>

<http://dna-view.com/profile.htm>

http://en.wikipedia.org/wiki/DNA_Fingerprinting

⁸ The "birthday paradox" may help to illustrate the point. The chance that a person chosen at random will share my birthday is approximately 1 in 365. But in a room of 23 people the chances of finding two people who share a birthday is actually greater than ½ (ie it is more probable than not). If the room contains twins, the chance is 100%. See: http://en.wikipedia.org/wiki/Birthday_paradox