Special Topic 3 - DNA Forensics

DNA forensics refers to the use of **DNA profiling** in Forensic science. These profiles can be used to determine identity, guilt, and innocence using biological samples. These samples can be matched against databases full of DNA profiles to aid the legal system.

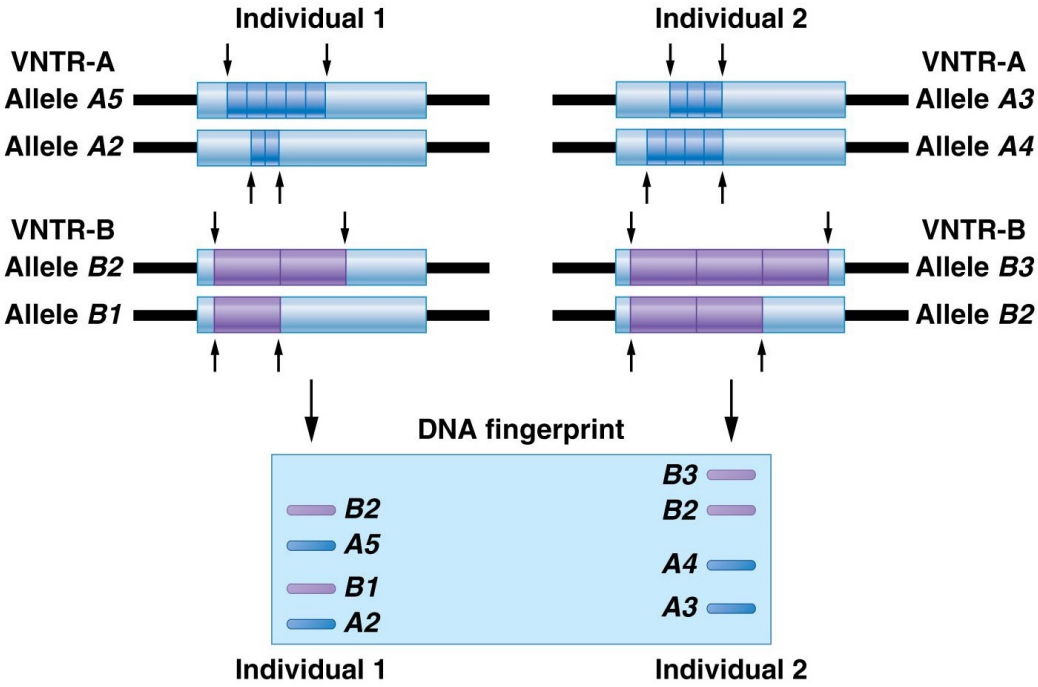
The persecution of criminals is, however, not the only use case of DNA profiling. Other areas of application include identifying victims of natural disasters, paternity and family relationship testing, identification of plant materials, and evolutionary studies.

Although being the gold standard of forensic identification, DNA profiling still has its flaws and controversial applications.

VNTR-Based DNA Fingerprinting

Variable Number of Tandem Repeats are 15 to 100 basepair long repeating regions within the noncoding parts of the genome. The number of repeats at a given VNTR locus varies from person to person, which results in lengths varying between 1 and 20 kilobases depending on the person. As many as 30 different possible alleles (repeat lengths) may be found at any VNTR in a population. Translation: examining 4 VNTR loci with 20 possible alleles each would result in 4^{20} = over 1 trillion possibilities

To create a personal profile of these VNTR regions DNA is extracted from a sample, after which a restriction enzyme is added which cleaves on either side of the VNTR region. This cleaved DNA is then separated by gel electrophoresis and subjected to Southern blot analysis.



VNTR loci where the arrows indicate cleavage sites. These then result in the DNA fingerprint.

A person may be identified if enough VNTR regions are analysed, which in practise means about five or six loci are examined.

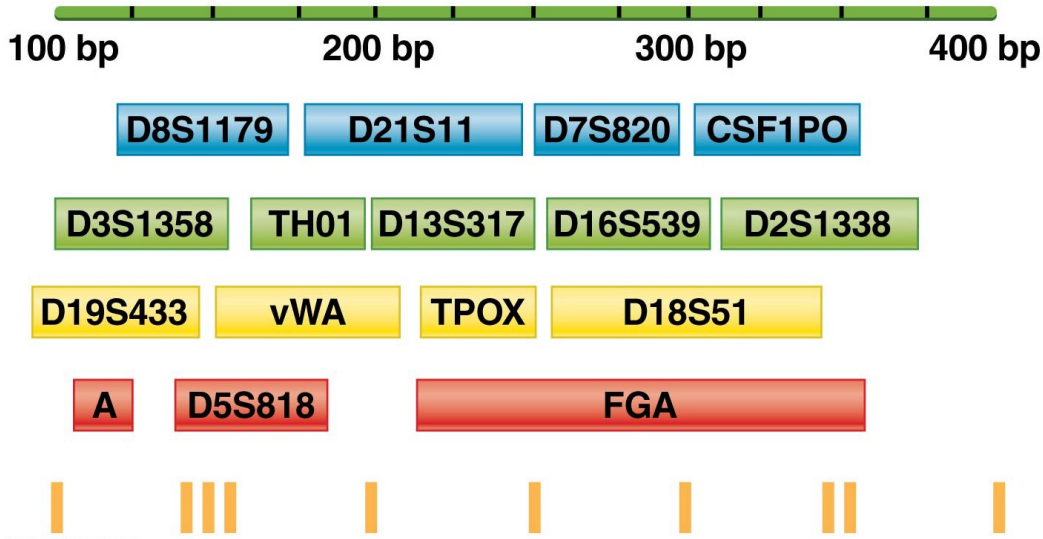
A drawback of this method is that it requires a relatively large sample of DNA, more than is usually found at a crime scene. Therefore it has mostly been replaced by autosomal STR.

Autosomal STR DNA Profiling

Autosomal Short Tandem Repeat regions are similar to VNTRs but are a lot shorter: between two and nine basepairs repeated 7 to 40 times. There are hundreds of STR loci present within the human genome but a subset of only 13 of these are used by US law enforcement, while 12 are used in most European countries.

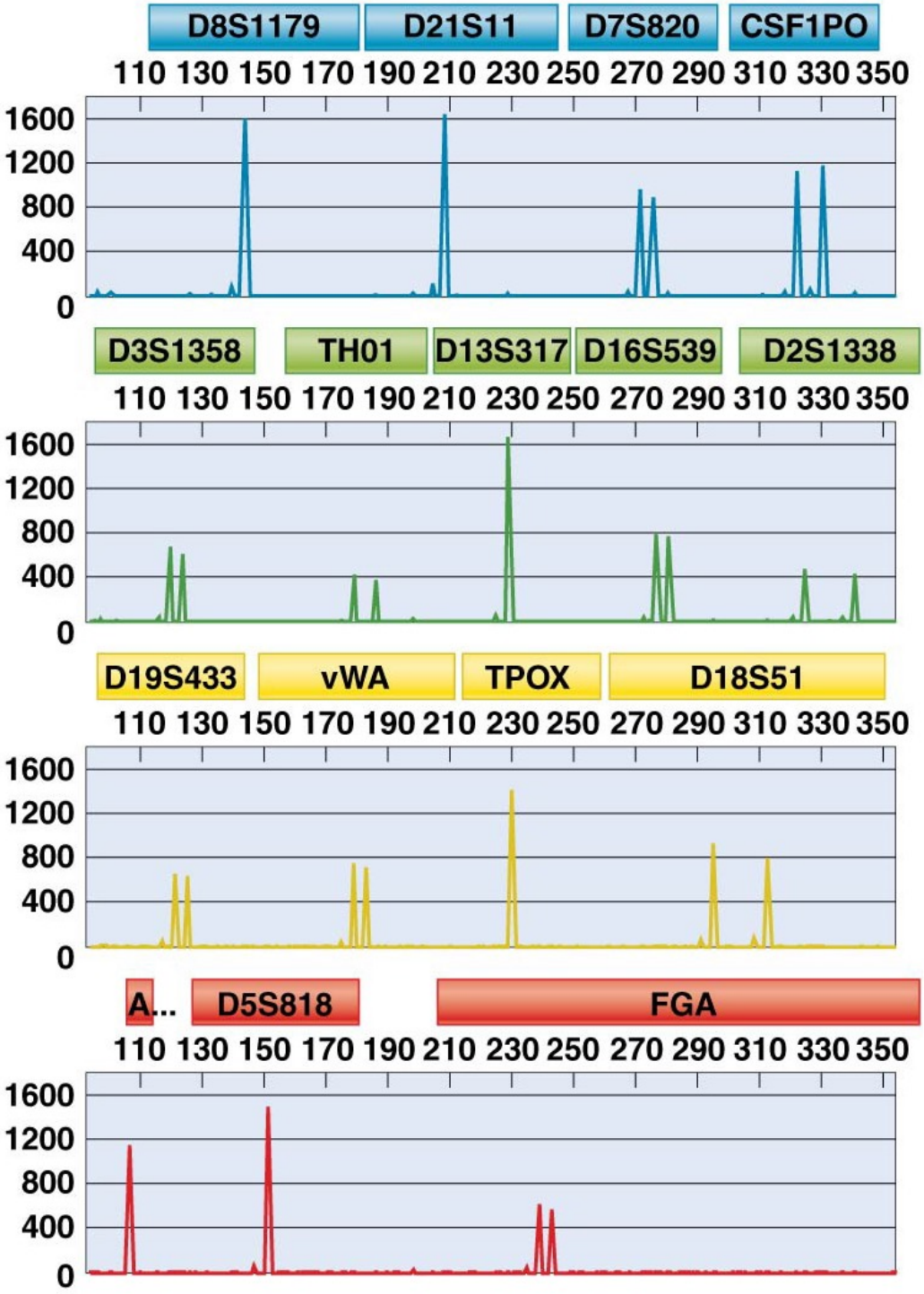
Identifying a person from STRs makes use of PCR for which commercially available kits are mostly used. With these kits each primer is tagged by a blue, green, yellow, or red fluorescent dye. Each set of primers (e.g., coloured tag) is designed for DNA amplification. After amplification, the DNA sample will contain a small amount of the original template DNA and a large amount of the

fluorescently tagged DNA. STRs of differing lengths are grouped together so that they can still be easily identified after amplification. For example: STR A & B are 20 bp long, while STR C & D are 30 bp long. STR A & C may be labeled green, while STR B & D may be labeled blue. This way the STRs with the same length (A & B and C & D) can be differentiated by their tag, while STRs within a certain tag can be differentiated by their length.



STRs of differing lengths grouped together by differently coloured tags.

The sizes of the amplified fragments are measured using **capillary electrophoresis**: A glass tube is filled with a polyacrylamide gel and the DNA sample is loaded onto the top of the tube. An electric current is passed through the tube which separates the fragments by their size. At the bottom of the tube a laser detects each fragment, with the smaller fragments moving faster and arriving first. This results in a unique selection of peaks for a given person, an example of which can be found in the graphs below.

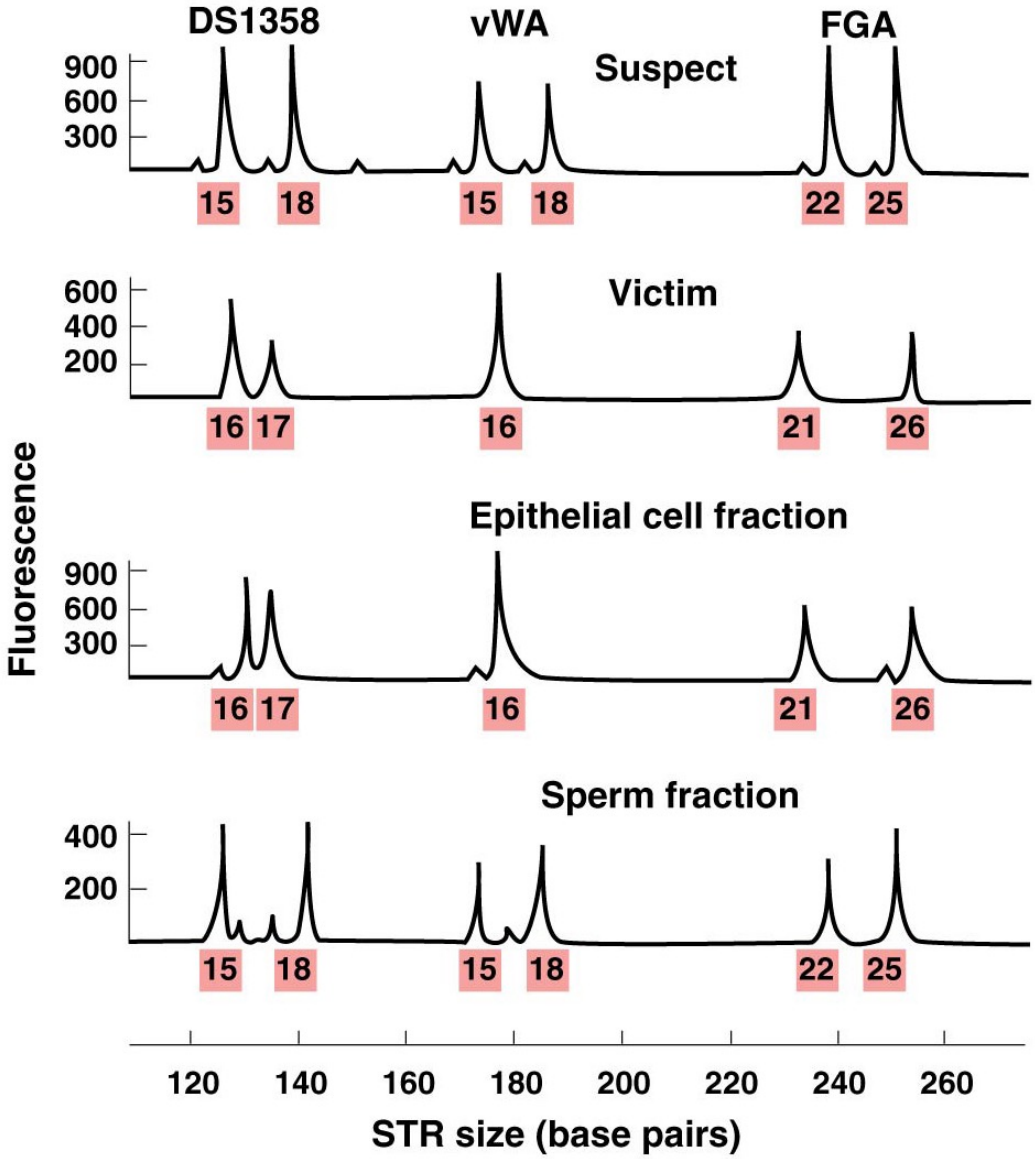


STRs separated by size (x-axis) and coloured tag.

The unique profile for a given person is expressed as the number of times the STR sequence is repeated.

Y-Chromosome STR Profiling

In some forensic applications it may be important to differentiate between two or more people in a mixed sample. Take for example a vaginal swap from a rape case, which may contain cells from both the male suspect as well as a female victim.



STR profiles of four different samples from a rape case.

In addition, some samples may contain material from multiple male suspects. In this case creating an STR profile of specifically the Y-chromosome may be useful. For this specific primers are used during PCA which do not amplify DNA on the X-chromosome.

This technique has one major drawback: Since the Y-chromosome is directly passed from father to son without recombination, this technique cannot differentiate someone from their father or male siblings. Even two apparently unrelated males may share the same profile if they share a distant male ancestor. Although this anomaly presents extra challenges for forensic analysis, it is actually quite useful for identifying missing persons when a male relative's DNA is available. It also allows researchers to trace paternal lineages more easily in genealogical studies.

Mitochondrial DNA Profiling

Mitochondrial DNA Profiling utilises the circular DNA found in mitochondria. The mitochondria are are passed from the human egg cell to the zygote during fertilisation, where the sperm cell contributes very few, if any, mitochondria. Therefore mtDNA profiling is similar to Y-chromosome profiling but it is passed through the maternal lineage. This mean that it also suffers from the same drawbacks as Y-chromosome profiling: it undergoes very little recombination which makes differentiation within a maternal lineage impossible.

mtDNA profiles are created by amplifying regions that show variability between unrelated individuals. Two commonly used regions for this are Hyper Variable Segment I and II (HVSI and HVSII). mtDNA is present in high numbers within cells, making it very useful in cases where small, old or degraded samples need to be analysed.

SNP Profiling

In case you didn't pass biology 1, **Single Nucleotide Polymorphisms** are single nucleotide changes between two DNA molecules. These include changes from one nucleotide to another as well as insertions and deletions. They occur randomly throughout both the genome and mtDNA creating millions of individual loci which may be used for profiling. However, as SNPs usually only have two alleles a lot more individual loci (50+) are to distinguish between individuals as efficiently as STRs.

SNP profiling is particularly well suited for analysing degraded samples since the theoretical size of DNA required for PCA is the size of the two primers plus one nucleotide (the SNP).

Interpreting DNA Profiles

If a suspect's DNA profile does not match an evidence sample or profile present in a DNA database, the conclusion is rather easy; the suspect is not the source of the sample. Things get a lot more complicated when there is a match though. If this is the case two conclusions are possible. Either the suspect's profile matches the evidence / database profile, or they came from two different people who share the same DNA profile by chance.
The **profile probability** or **random match probability method** gives a numerical probability that a person chosen at random from a population would share the same DNA profile as the evidence or suspect profiles. The table below demonstrates how to perform this calculation.

ST TABLE 3.2 **A Profile Probability Calculation Based on Analysis of Five STR Loci**

STR Locus	Alleles from Profile	Allele Frequency from Population Database*	Genotype Frequency Calculation
D5S818	11	0.361	$2pq = 2 \times 0.361 \times 0.141 = 0.102$
	13	0.141	
TPOX	11	0.243	$p^2 = 0.243 \times 0.243 = 0.059$
	11	0.243	
D8S1179	13	0.305	$2pq = 2 \times 0.305 \times 0.031 = 0.019$
	16	0.031	
CSF1PO	10	0.217	$p^2 = 0.217 \times 0.217 \times 0.047$
	10	0.217	
D19S433	13	0.253	$2pq = 2 \times 0.253 \times 0.369 = 0.187$
	14	0.369	
Genotype frequency from this 5-locus profile = $0.102 \times 0.059 \times 0.019 \times 0.047 \times 0.187 = 0.0000009 = 9 \times 10^{-7}$			

*A U.S. Caucasian population database (Butler, J.M., et al. 2003. *J. Forensic Sci.* 48: 908–911).

*Calculation of profile probability by multiplying the different genotype frequency calculations together, known as the **product rule**.*

In the case of the five-locus profile above 9 out of 10 million (~ 1 in a million) chosen at random would share this DNA profile.

Uniqueness of DNA Profiles

As the number of analysed loci is increased the probability of a random match decreases. Therefore, if enough loci were analysed one could be *almost* certain that a given DNA profile is unique. Currently North American law enforcement use a core set of 13 STR loci to generate DNA profiles. The frequency of this profile would be 1 in 10 billion, so unique enough to identify every living person on earth! But as always, there are some exceptions to this rule.

The first is in the case of identical twins. Since they share the same DNA they will also produce an identical DNA profile.
Then there is the issue of small interrelated groups. For example, siblings can share one allele at any DNA locus in about 50 percent of cases and can share both alleles in a locus in about 25 percent of cases.

The allele frequencies and calculations that were described above assume that the population is large and has little relatedness or inbreeding. If some of these assumptions aren't valid in a certain situation (as with siblings for example) the profile probabilities must be adjusted to take this into account.

The Prosecutor's Fallacy

Equating guilt with the numerical probability of one piece of evidence is known as **the prosecutor's fallacy**. An example would be a statement like "*the suspect must be guilty given that the chance of a random match to the crime scene sample is 1 in 10 billion—greater than the population of the planet.*"

A match between a suspect's DNA profile and evidence does not necessarily prove guilt for reasons such as contamination, human error, or deliberate tampering. In the same vein, a DNA profile that does not provide a match doesn't equate innocence either. A sample might not match a suspect's DNA, but they might still be involved in some other way. DNA profiles must therefore be considered in the context of all the evidence in a case.

DNA Profile Databases

Combined DNA Index System (CODIS) is a system of databases and analysis tools maintained by the FBI to assist in criminal cases and missing person reports. It consists of two main databases: the **Convicted Offender Database** containing DNA profiles of...well...convicted offenders, and the **Forensic Database** containing profiles of crime scene evidence.

Technical & Ethical Issues Surrounding DNA Profiling

Although DNA profiling is a sensitive and accurate technique, it isn't without its limitations. These limitations can be decided into: **general limitations** of DNA profiling, **technical limitations**, and **ethical issues**.

Limitations of DNA profiling

DNA profiling is limited by two main factors:

- The fact that most criminal cases have **no DNA evidence for analysis**, or have DNA evidence which is not informative to the case.
- The fact that a lot of potentially valuable **DNA evidence remains unprocessed and backlogged**.

Technical limitations

The first serious technical limitation is that of **human error**. There are cases of DNA samples being accidentally swapped and getting innocent people convicted.
Secondly, a crime scene may be **contaminated** by biological material which was indirectly introduced to the site.
Last but not least, crime scene evidence is often severely **degraded**, producing only partial DNA profiles.

Ethical issues

A more disturbing issue than the ones listed above is **deliberate tampering**. There have been cases in which criminals have introduced biological material to crime scenes in an attempt to affect DNA profiles. In addition to this it is possible to **manufacture artificial DNA segments** that match STR loci of a person's DNA profile, mix this with bodily fluids, deposit this sample on crime scene items, and producing perfect STR profiles using forensic analysis based on these artificial samples. It may therefore be necessary to produce methods for detecting synthetic or cloned DNA based on epigenetic markers such as methylation.

Another ethical issue surrounding DNA profiles is the **storage and use of these profiles** by certain organisations. Who should store this data and who should have access under which circumstances? Should a suspect's knowledge or consent be required before storing their DNA profiles in a database?

Additionally there is the issue of **partially matching DNA profiles**. A profile gathered from crime scene evidence may only partially match a suspect's profile. Is it ethical to then perform **familial DNA testing** by gathering DNA

from closely related individuals?

Lastly there is a method known as **DNA phenotyping** which is the practice of accurately predicting a suspects phenotype (eye colour, age, height, racial ancestry, etc.) based on their DNA profile. Could a suspect be identified or even convicted by matching their DNA profile to their phenotype? And would this methodology be ethically sound and scientifically valid?

All of the problems described are things that should be carefully considered as DNA profiling becomes more powerful, sophisticated, and prevalent.