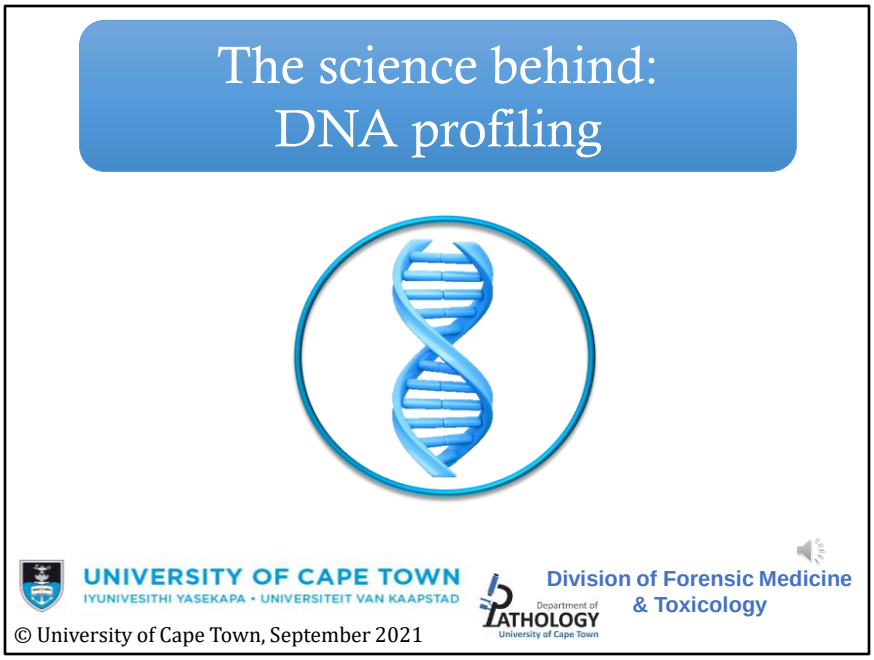
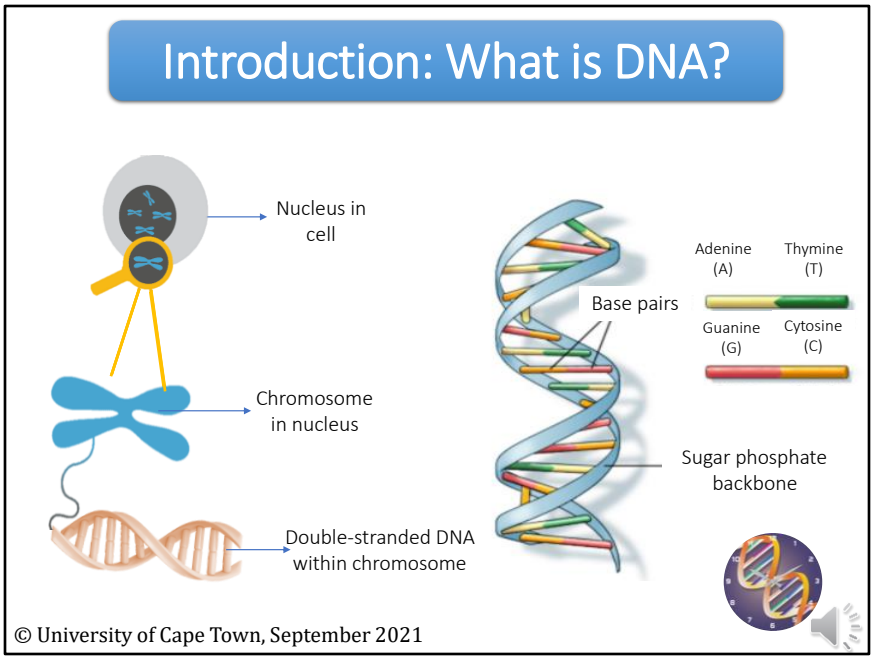
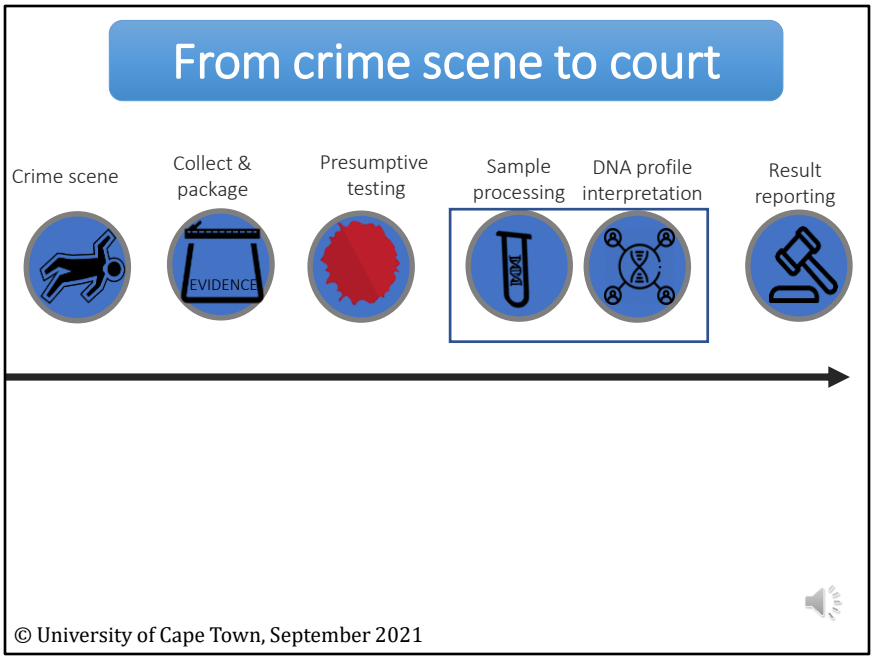


The science behind DNA profiling.

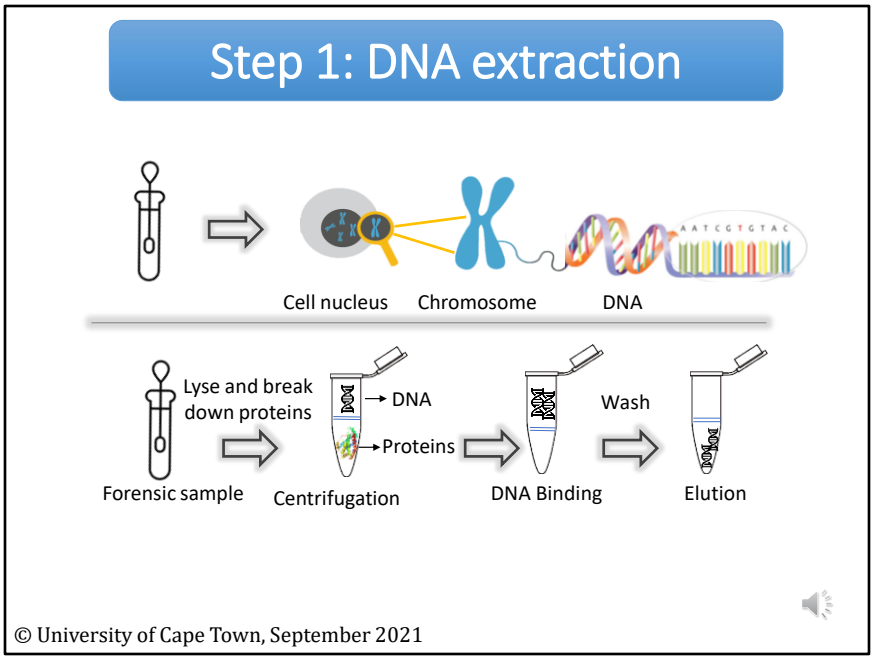
DNA is the molecule that makes you who you are. It is made of long strings of different compounds or base pairs in a particular order. The way DNA sequences are repeated is highly discriminating between people. So, if you have a DNA sample, that would be ideal for identifying an individual. DNA profiling, sometimes called DNA fingerprinting, DNA typing, DNA identification, is rapidly becoming central to medicolegal investigations. This lecture provides an explanation of the science behind the use of DNA in a medicolegal context. The process begins with the extraction of DNA from a piece of evidence or perhaps a vial of blood or a buccal swab. This is then quantified to determine how much usable DNA is present. This is then amplified and measured, and used to generate a DNA profile that can be used to identify an individual.



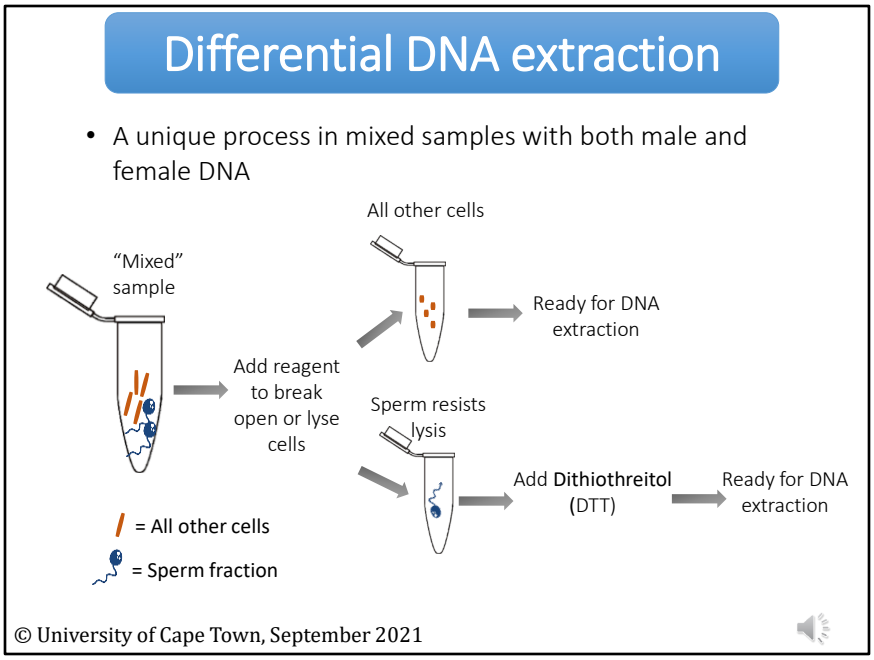
DNA , which stands for Deoxyribonucleic acid is a molecule that is found inside cells in your body. Everybody except for identical twins have their own sequence of DNA that contains information that determines your growth, appearance, sex and reflects your ancestry. DNA molecules are tightly packed into thread-like structures called chromosomes. The information in DNA is stored as a code made up of four chemical bases: adenine (A), guanine (G), cytosine (C), and thymine (T). Only certain DNA bases can pair up with each other, A with T and C with G, to form units called base pairs. Everyone's DNA has certain sections with repeating patterns of base pairs, but the number of times the pattern repeats itself varies from person to person. These repeated sections are called short tandem repeats and are used to generate a DNA profile. We will focus on short tandem repeats later in this lecture.



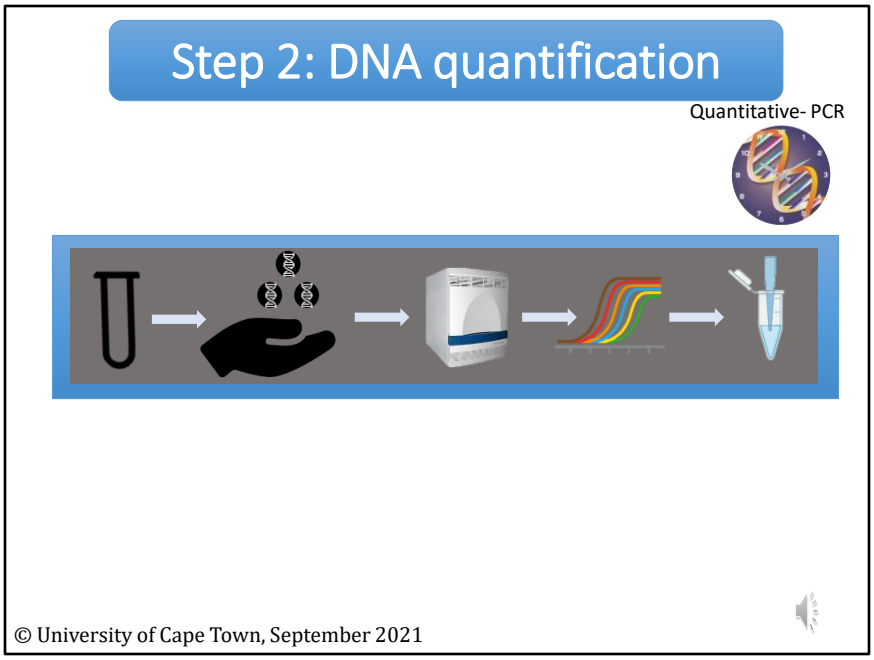
So how is DNA used in a medicolegal context? If someone deposits traces of their blood, hair, saliva or semen at a crime scene, crime scene investigators collect, package and transport these samples to the laboratory. Samples of blood, different cells and tissues collected from living and deceased individuals can also be sent to the laboratory for DNA analyses. In samples with an unknown stain, the first step is to conduct a presumptive test to determine whether the sample is biological. This is performed on items with unknown stains to ensure that the sample is biological before the DNA is extracted, processed and analyzed. Based on the appearance of the stain and the case history a presumptive test for blood, semen, or urine can be conducted. See our lecture on body fluid identification to find out more about these tests. Because the protocol used to extract DNA is slightly different for different tissue types like sperm, blood, hair and saliva, by determining what tissue type is present, the forensic geneticist can ensure they use the most optimal extraction processed.



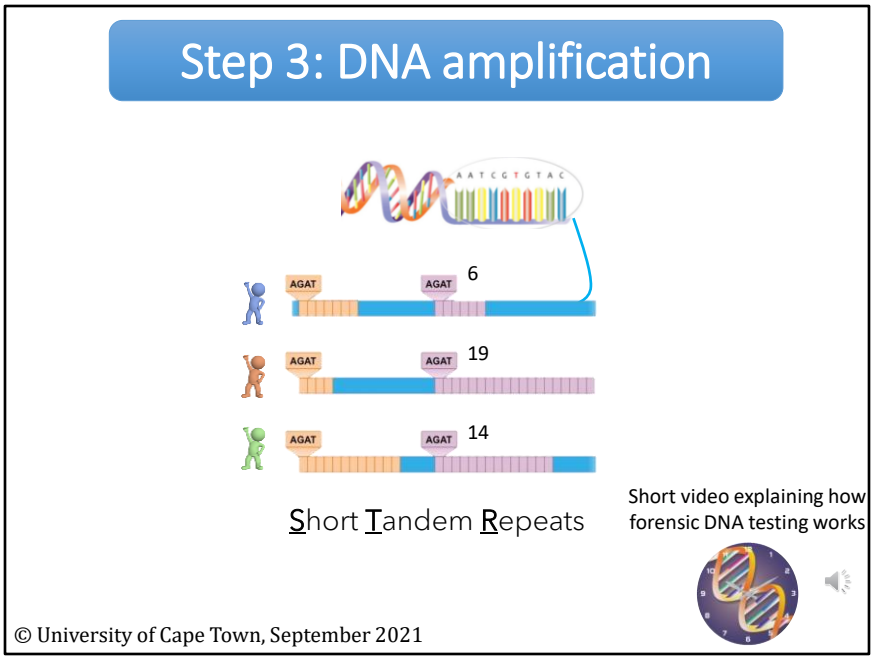
The first step in the DNA workflow is DNA extraction which is a routine procedure used to isolate the DNA from the nucleus of a cell. In order to perform DNA testing, the DNA needs to be released from the cell, and preferably, isolated from the other cellular components. A detergent is used to break open the cell membrane which releases all the cellular contents. This step is called lysis. Chemicals are then added which break down the proteins in the cell, because they are not needed for DNA analyses, and if left untreated, can even start to degrade the DNA. The cell components and broken-down proteins then need to be separated and essentially removed from the DNA. The solution we are left with contains with pure DNA. This step is referred to as purification. There are many ways to achieve this, such as relying on solubility of DNA in some solutions but not others. Or by centrifugation, which uses a rotational force to separate components in a mixture based on size, weight or density. Because proteins are denser than DNA, they sink to the bottom of the tube during centrifugation. The lighter DNA molecules sit on top of the mixture, and this is carefully removed for profiling.



There are unique considerations when evidence is from a suspected sexual assault. Evidence from sexual assault cases usually like vaginal swabs, anal swabs, or cuttings from underwear contains an cells and DNA from both the victim and the perpetrator. In the male-on-female sexual assaults this results in an imbalance of female to male DNA. For this reason, a differential DNA extraction is performed on samples from sexual assaults that are likely to contain both male and female DNA. This process is simply a regular DNA extraction, but with a modification that helps to isolate DNA from sperm, and separate this from the other cells. This is possible because sperm cells are more difficult to lyse, or break open. During the first normal lysis step, DNA is released from the non-sperm cells only. This is then separated from the sperm which still contains DNA, a chemical called Dithiothreitol or DTT is then added to break open the sperm cells and release the DNA. The DNA extracted from the sperm cells is purified separately from the DNA extracted from the females epithelial cells.

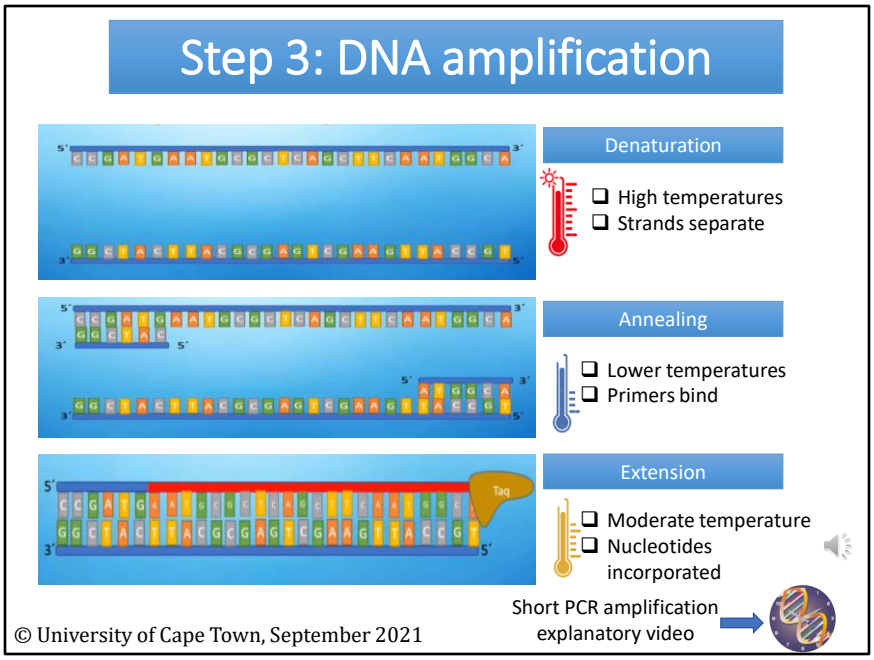


The next step after extracting DNA is assessing how much usable human DNA is present. This is done to determine whether there is enough DNA for the STR profiling. If there is too much DNA, the sample is diluted with sterile water, and if there is too little DNA, another round of PCR is used to make more copies, or we will use lower volumes for the reaction. One of the most used methods of quantification is qPCR which stands for quantitative real time polymerase chain reaction. In this process, the concentration of DNA in a sample is measured against a DNA standard of known concentration. This is done through a process of amplification, where areas in the DNA are multiplied. Some of the qPCR kits have a targets on the Y Chromosome, which can indicate if there is male DNA present in the sample, already at this stage.



The most routinely used technique for DNA profiling relies on differences in target areas of DNA called short tandem repeats (or STRs). STRs or short tandem repeats are repeated sequences of DNA base pairs. Everyone's DNA has certain sections with repeating patterns of base pairs, but the number of times the pattern repeats itself varies from person to person. An STR profile usually contains information for 13 or more STR's. The number of repeats within an STR is known as an allele.

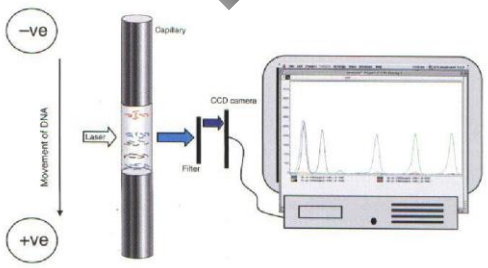
Here is an example below, of the AGAT nucleotide sequence shown in purple. As you can see all three individuals have this sequence in a specific region of their DNA. But if you look closely, the number of times this is repeated is different between individuals. The purple AGAT unit is repeated 6 times in individual 1, 19 times in individual 2 and 14 times in individual 3. This means individual 1 has allele 6 at this STR location or locus, individual 2 has allele 19 at this STR locus and so forth. The combination of the different alleles at the different loci is what makes is possible to identify an individual from their DNA. Each person inherits two alleles per STR, one from their mother and one from their father.



Often only small amounts of DNA are available for forensic analysis so the STRs are copied or amplified many times using the polymerase chain reaction (PCR). A special machine called a PCR machine is used to conduct this and it requires exposing the DNA to different temperatures to facilitate the copying process. How does PCR work? During PCR, the double-stranded DNA is first denatured or separated into single-strands by exposure to a very high temperature. Primers, which are molecule containing short sequences of DNA that are complementary to the regions next to STRs are added to the reaction. These primers will bind to the single stranded DNA at a lower temperature, in a process is called annealing. An enzyme called as DNA polymerase or Taq that is added to the reaction is activated to make copies of the STR regions. After the first round of PCR, the number of copies doubles in the regions of interest. This process then happens over and over again until millions of copies of the STRs are made.

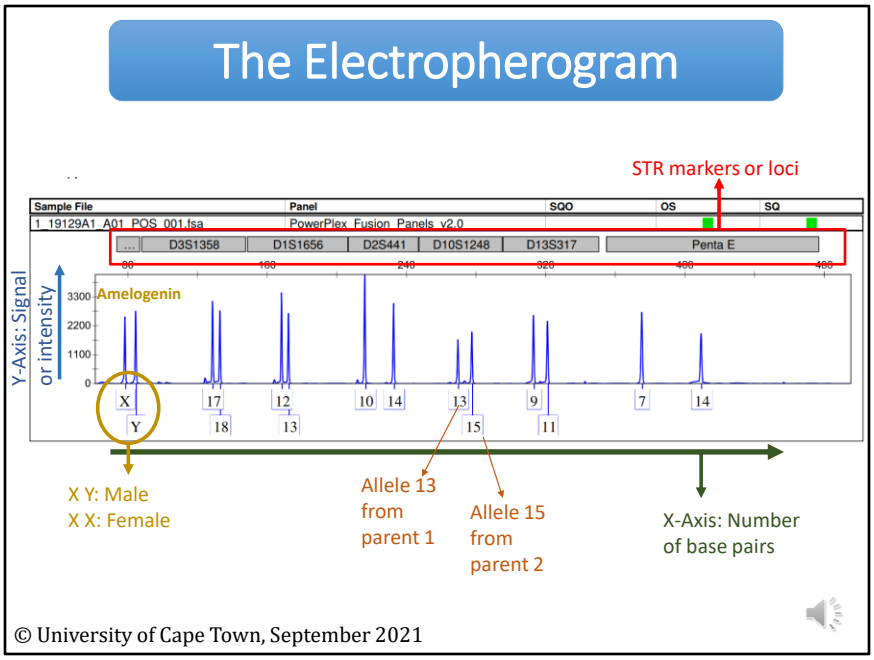
Step 4: Generating the DNA profile

Amplification product is injected into a capillary system

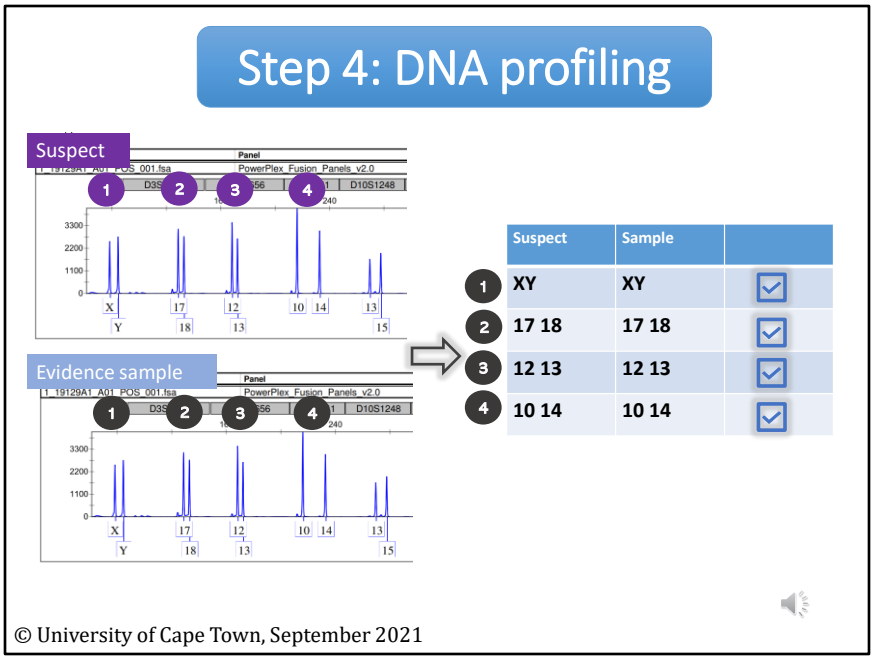


Capillary Electrophoresis

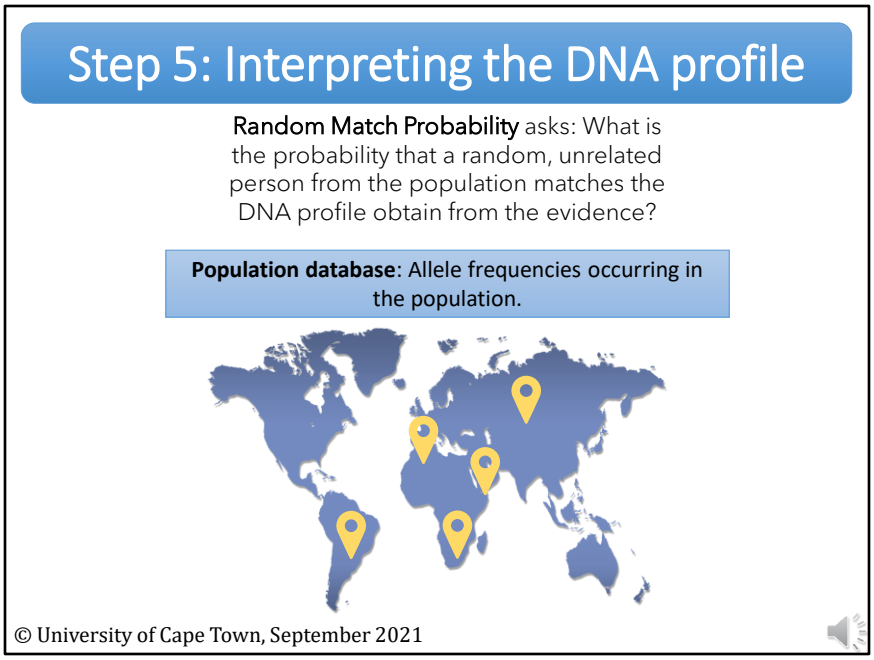
Once the STRs have been amplified, the size or the length of each STR is determined to generate a DNA profile. This is measured with a genetic analyzer in a process called capillary electrophoresis. A genetic analyzer separates the amplified regions of DNA according to their sizes. Different STR sizes are detected and represented by peaks on a graph called an electropherogram. Each individual has a different combination of the sizes of the STRs in their DNA. Or a different combination of 2 alleles at each locus. This is what makes up the DNA profile



Here is a example of an electropherogram. Here the horizontal line, indicates the number of base pairs for each allele. The vertical line represents the signal or intensity of the fragments detected by the capillary electrophoresis system. The STR markers are shown in grey at the top of each electropherogram. The first marker shown in yellow is Amelogenin. Amelogenin is a marker used to identify the sex of the individual, females have two X alleles and males have an X and a Y allele. In this example, we the electropherogram registers an X and a Y allele, which is consistent with this individual being a male. The next marker is called D3S1358, and at this STR marker alleles 17 and 18 are present. And if you look at the next marker, D1S1656 and alleles 10 and 14 are present. The combination of the different alleles in each individual is highly discriminating and can be used for identification.

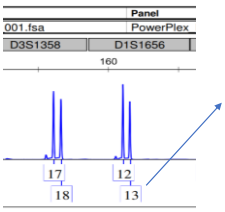


Suppose we have obtained a sample of DNA from a suspect. We want to determine whether the sample we obtained from a piece of evidence matches that of the suspect or not. The DNA is extracted, purified, quantified, amplified, and an STR profile is obtained. A forensic geneticist would then analyse the profile from the suspect and sample to determine whether they match. As you can see, at the first locus (labelled 1), we see that both profiles have the same alleles. The sample from the suspect has an X and Y profile at the first locus, this indicates male sex. At locus 2, both profiles have alleles 17 and 18, and matches are also seen for locus 3 and 4. If all the alleles on both profiles match, this may suggest that the DNA on the evidence is from the suspect. But, how do we know if this match is true? How do we know that another individual in the population doesn't have the same combination of alleles as the DNA profile in question?



To answer these questions, we need to determine what is the probability is that a random, unrelated person from the population matches the DNA profile obtained from the evidence. This is called random match probability. To calculate this probability, we use something called a population database. A population database, tells us how frequently an allele occurs in a population.

DNA profile frequency

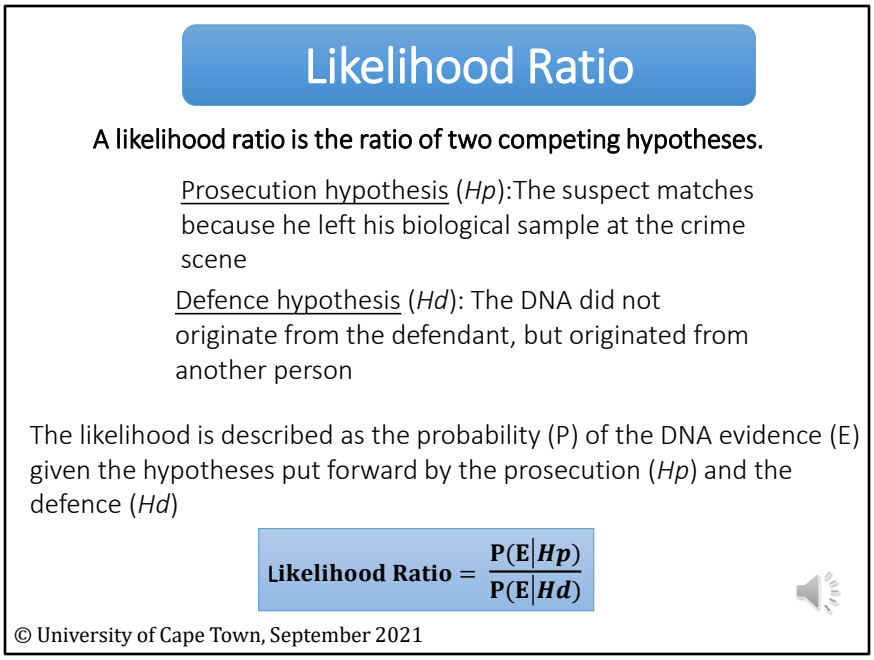


How **often** does this combination of alleles occur in our population? (i.e. what is the **frequency** of this combination of alleles?)

- The **DNA profile frequency** (or the frequency of each combination of alleles at each marker) = **1 x 10^{-7}.**
- The probability of this DNA profile belonging to a random, unrelated person in the population is:

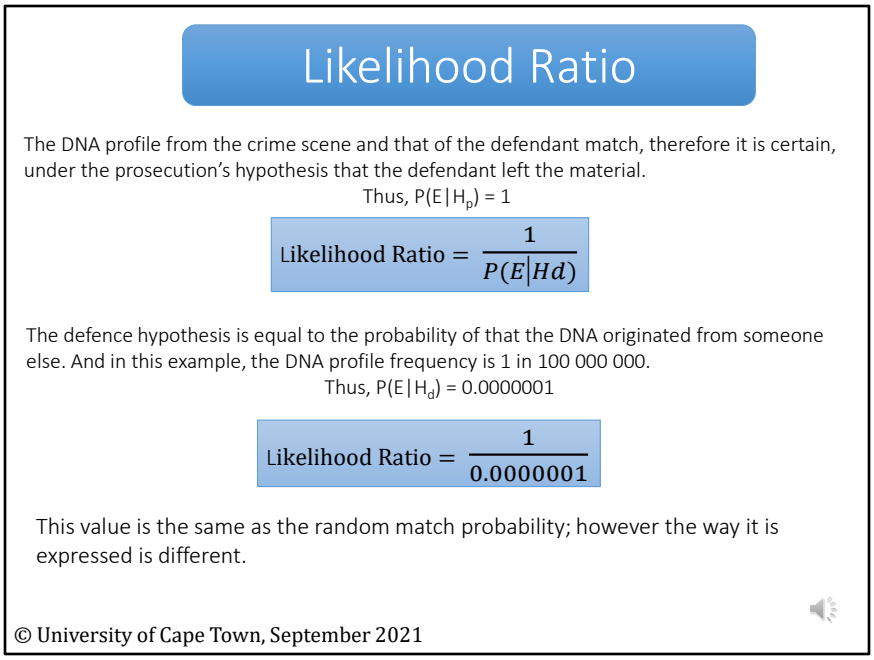
1 in 10 000 000

A DNA profile frequency is determined by looking up the how often each combination of alleles occurs in the population. Forensic geneticists look at the combination of alleles at each marker, the combined result is reported, in addition to the probability that that specific DNA profile occurs by chance in the population. If the combined frequency of the DNA profile is calculated to be 1x 10^-7, then that means that the probability that the DNA profile belongs to an unrelated individual in the population, is 1 in 10 million. This means we can be more confident that the DNA profile does belong to the person in question and not a random individual in the population.



The random match probability approach has the advantage of being simple and attractive to the court. Representing these frequencies does, however, pose some problems.

- For example, quoting the DNA match probability of 1 in a billion in a country, where there are less than a billion people, is problematic. One may ask which population is being referred to. This is addressed by using the likelihood ratio.
- A likelihood ratio is the ratio of two competing hypotheses. In a criminal investigation it is the ratio of the prosecution hypothesis (*Hp*) and the defence hypothesis (*H*d). In the case of a DNA profile match, two possible hypotheses are:
- Prosecution hypothesis (*Hp*): The suspect matches because he left his biological sample at the crime scene, or
- Defence hypothesis (*Hd*): The DNA did not originate from the defendant, but originated from another person.
- The likelihood is described as the probability (P) of the DNA evidence (E); given the hypotheses put forward by the prosecution (*Hp*) and the defense (*Hd*)



- The DNA profile of the defendant and the evidence from the crime scene match, therefore it is certain under the prosecution's hypothesis that the defendant left the material, thus P(E|Hp) = 1
- The defence hypothesis is equal to the probability that the DNA originated from someone else. And in this example, the DNA profile frequency is 1 in a 10 million.
- This value is the same as the random match probability; however the way it is expressed is different.

This means that the greater the likelihood ratio, the stronger is the evidence in favor of the hypothesis corresponding to the numerator, that the source of the DNA from the evidence sample and the suspect are the same individual.
Once the DNA profiles have been analysed and interpreted, the results are written into an affidavit and reported in court.

Voice over: Donna-Lee Martin

Slide design: Donna-Lee Martin, Dr Laura Heathfield and Elizabeth Dinkele

Slides provided by the Division of Forensic Medicine and Toxicology, Faculty of Health Sciences, University of Cape Town

Contact person: Dr Laura Heathfield
Laura.Heathfield@uct.ac.za