DNA Analysis: A Brief Introduction

By Dr. Maher "Max" Noureddine, PhD, Founder of ForensiGen (http://www.forensigen.com); and Susan H. Johnson, MPH, CCCE Program Coordinator

DNA, or Deoxyribonucleic acid, contains all of the instructions that determine an individual's genetic characteristics – from type of organism (human versus amoeba), to gender, hair color, and eye color. DNA is present in all cells in the body except for red blood cells. Each person inherits two copies of DNA, one copy from each parent. The two copies in each cell make up what is known as the "genome."

All DNA contains four nucleotide bases – adenine, cytosine, guanine, and thymine (A, C, G, T). Approximately 6.3 billion genetic base pairs are present in DNA-containing human cells, in which C pairs with G, and A pairs with T. These pairs of bases combine to form a structure resembling a spiral staircase (the DNA helix).

The DNA of all humans is almost identical; nearly 99.9% of the A's, C's, G's and T's are in the exact same order. At numerous locations in the genome, however, these base pairs are known to vary among different individuals, giving forensic scientists the ability to distinguish between the DNA of different people. The variable regions that forensic scientists focus on are called STRs (for Short Tandem Repeats), and their variability is called "polymorphism." STRs are also known as "markers" because they reside in specific locations in the genome known as "loci" ("loci" is plural for "locus").

The order and varying combinations of the nucleotide bases determine the organism's genetic characteristics, from whether an organism is a human or a hamster, to whether a person has blue eyes or blond hair, to whether a person is at higher risk of certain diseases. We also have additional base pairs that do not seem to "code" for anything, or base pairs with functions that are not yet identified. Many STRs are located in regions with no known function. These STRs, however, are very useful in forensic investigations because they are highly variable between people. This variability allows forensic scientists to distinguish between DNA samples originating from different individuals.

Typically, forensic scientists examine 13 – 15 specific markers in the course of determining someone's genetic profile. Genetic profiles from tested individuals are kept in the CODIS (Combined DNA Index System) database. This searchable database can be accessed by forensic scientists and law enforcement agents when trying to match a DNA sample to a specific individual. More than 12 million DNA profiles are contained in this database; the majority comes from DNA samples collected from convicted offenders and/or arrestees. Other profiles are generated from evidentiary items, or collected from unidentified human remains and relatives of missing individuals.

When examining an STR marker (or locus), the goal of the forensic scientist is to determine the length of that STR marker. The length of an STR marker varies among individuals. A variation is known as an "allele." The possible types of alleles are labeled by a number (5, 6, 9, 11, etc.) that signifies the length of the STR marker: "5" means that the STR has 5 repeating units, "6" means that the STR has 6 repeating units, and so on. An allele 5 is shorter than allele 6 by one repeat unit, a difference that can be detected through laboratory analysis.

Each person inherits one allele from each parent. If the person inherits the same allele from both parents (for instance, two "5s"), the person is said to be "homozygous" at that marker. If the person inherits a different allele from each parent (for instance, a "5" from mom and an "8" from dad), the person is "heterozygous" at that marker.

When a forensic expert tests an evidence sample to try to confirm a DNA match, the sample undergoes a process by which the key markers are targeted and copied in order to increase the amount of DNA available for analysis. This copying process is known as "amplification." After a large number of copies are made, the DNA fragments are separated by size using a process called "electrophoresis." The resulting information is then displayed in a graph called an "electropherogram," which has a series of sharp peaks at specific spots associated with each locus.

The height of each peak indicates the relative quantity of DNA from which the peak was generated. For instance, if an individual inherits different alleles from his/her parents (say 5 and 7), the peaks representing those alleles will be separated from each other and will have similar heights. If an individual inherits the same allele from both parents (say 7 from mom and 7 from dad), the one peak representing that allele will appear relatively higher on the electropherogram because of an additive effect (twice the amount of starting DNA for that allele). Furthermore, this relative quantity can be used by the forensic scientist to determine the possible number of contributors to a DNA sample.

By comparing the size and location of the peaks on the DNA evidence sample electropherogram to the electropherogram of a sample from a victim or suspect, the forensic scientist can determine whether there is a match or a non-match between victim/suspect and the evidence sample. If a "match" is determined, then the forensic scientist will calculate the probability of this match occurring by random chance alone (a statistic called Random Match Probability or RMP). If the odds of a random chance are small, then the likelihood is high that the DNA sample originates from the victim or suspect in question. It is theoretically possible that no two individuals will have the same DNA profile except for identical (monozygotic) twins.

The results from the DNA analysis can be displayed in an "allele-call table." This table lists the markers being analyzed (which have names such as "D8S1179) and the number of the allele type(s) found at each locus. By convention, if an individual inherited the same allele from each parent at a particular locus, only one allele will be listed, whereas if the individual is heterozygous at a certain marker, both alleles will be listed.

As with any scientific test, the final results can be affected by several factors, including the quality of the DNA sample, contamination of the sample, the presence of a single DNA source or multiple sources in the sample, the proficiency of the person collecting or processing the sample, the accuracy of the machinery used to process the sample, inadvertent errors introduced during sample copying, and the expertise of the person analyzing the results.

Let us now look at sample electropherograms developed by Dr. Max Noureddine of Forensigen, who helped to write the DNA analysis section in our competition case. Below, we have electropherograms from three fictional samples: DNA from an Evidence sample collected from a deceased victim after a struggle, such as mixed blood spatter or fingernail scrapings (pages 1-2 of

the electropherograms); the DNA sample from the suspected criminal (page 3); and a DNA sample solely from the deceased victim (p. 4).

To understand how a forensic analyst would determine whether the Evidence Sample matches the sample from the suspected criminal, let us look at the alleles for marker D8S1179 on all three samples. This marker is the first graph at the top left of each electropherogram, labeled "D8S1179" in the thin gray bar above the graph itself.

The Evidence Sample graph displays alleles 11 (relative quantity: 2302), 13 (relative quantity: 1290) and 14 (relative quantity: 873). Thus, we know already that the sample contains DNA from more than one person, as one individual could only have two alleles at a specific locus (one from each parent, if the parents' alleles differ), although there are known (rare) exceptions to this rule. Also, you'll notice that the relative quantity of allele 11 is significantly greater than that of alleles 13 and 14. The absolute amount of each is less important than the proportional amounts of the alleles relative to each other. Before reading on, think about why these amounts might differ. Which alleles do you expect to find in the other two samples?

The second electropherogram is based on the reference DNA sample from the Suspect. The data is consistent with the fact that this electropherogram contains DNA from only one individual—all markers show either one allele (homozygous) or two alleles (heterozygous). Looking at the results for the first marker (D8S1179), we see that allele 11 (relative quantity: 2124) and allele 14 (relative quantity: 1910) have similar relative quantities. Relative quantities will no longer be deemed similar if there is a 50% or greater difference between the relative quantity values from two alleles at the same marker. Look at the remaining three loci on the top line of the Suspect's electropherogram. Each of the three loci only contains data from one allele, which means that the Suspect inherited the same allele from both parents in all three loci (D21S11, D7S820, and CSF1PO). You can see this data in both the allele-call table as well as the electropherograms. Look at the remaining "heterozygous" markers, do you see any marker with alleles that are significantly different in terms of relative quantity?

Finally, the third electropherogram is based on the DNA from the deceased victim. The results for the first marker show alleles 11 (quantity: 719) and 13 (quantity: 990) – two alleles in roughly equal amounts. Again, each locus in the entire electropherogram has at most two alleles, revealing that the DNA sample comes from only one individual. If you look on the second line at the locus named "TH01," you can see that the victim was homozygous at this marker on the chromosome. Only one allele, "6," is present, indicating that the victim received allele "6" from both parents.

Let's use what we learned above by comparing our three DNA samples. At marker D8S1179: the Suspect has alleles 11 and 14; the victim has alleles 11 and 13; and the Evidence Sample has alleles 11, 13, and 14 with much greater relative quantity of allele 11 than of allele 13 or allele 14. If you do a similar comparison of every other marker on the electropherograms, you will see that nearly every marker displays alleles that correspond to a combination of the alleles from the suspected criminal and the victim. If the Suspect and the victim share alleles, you will likely see the effect of this sharing. Shared alleles will manifest in higher peaks.

The only exceptions are marker CSF1PO (first row, far right) and marker D13S317 (second row, middle), both of which only contain alleles from the victim. In copying DNA samples for analysis, sometimes strands of DNA do not copy in sufficient quantity to be fully detected or properly identified; such alleles are known as "dropouts." In the electropherograms for this case, an expert could conclude that the Suspect's DNA "dropped out" at those two markers.

Thus, 13 of the 15 markers (not counting the one for gender) contain alleles from both the Suspect and the victim. These results could lead a forensic analyst to conclude that the suspect is a "match"; calculations too complicated to discuss here would allow the forensic specialist to calculate an RMP for that match. (Note: all three electropherograms reveal individuals with the same gender; can you tell what that gender is?)

To explore this topic further, check out the other links on the CCCE website. DNA forensic analysis is a fascinating field! We hope you enjoy this opportunity to learn more about it as you participate in our mock trial program.

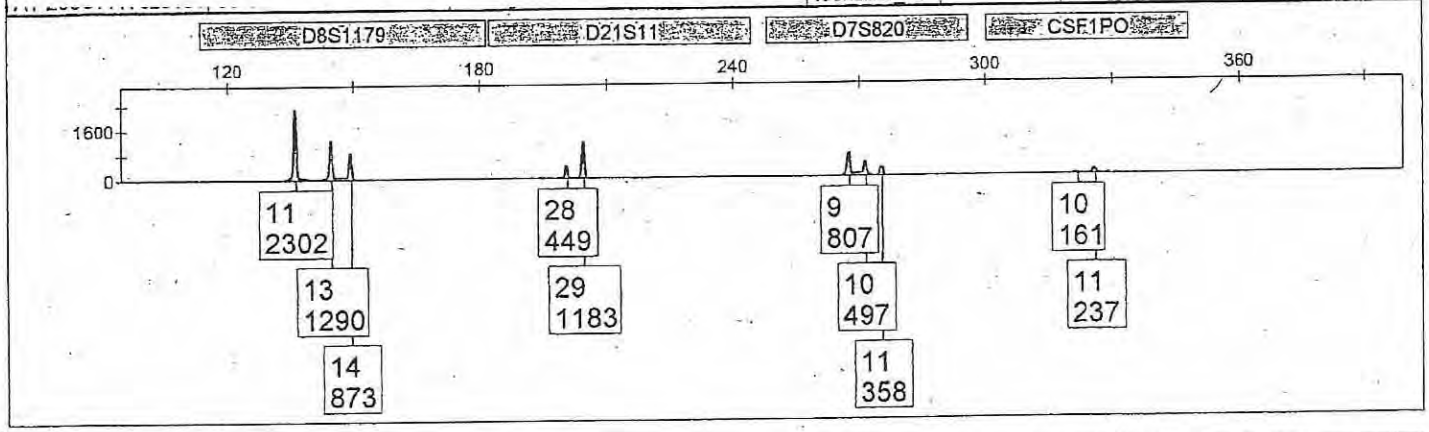
Applied Biosystems

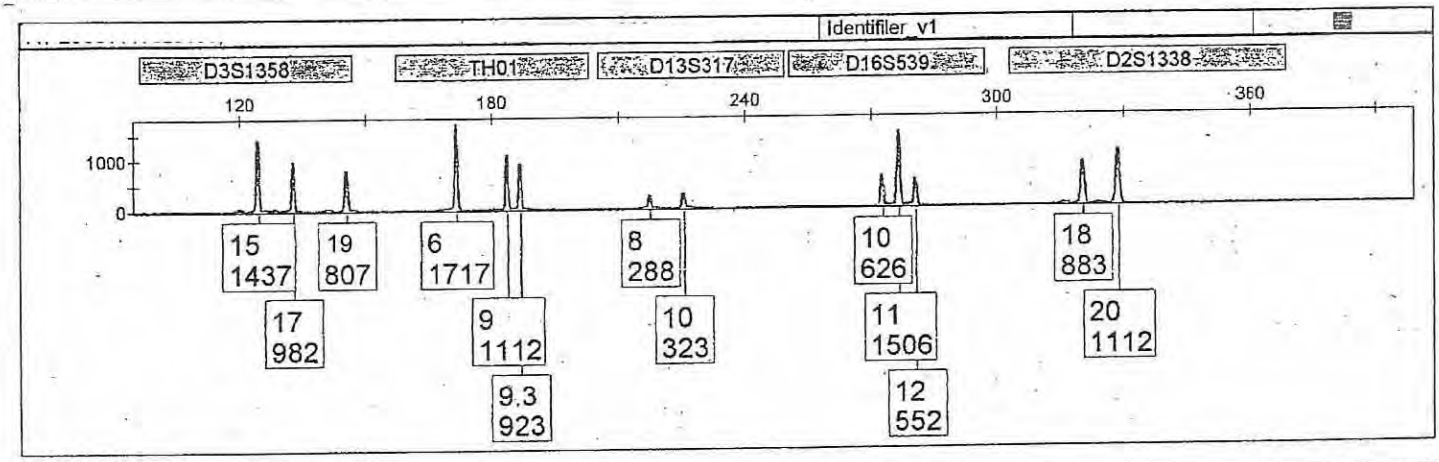
GeneMapper ID v3.2.1

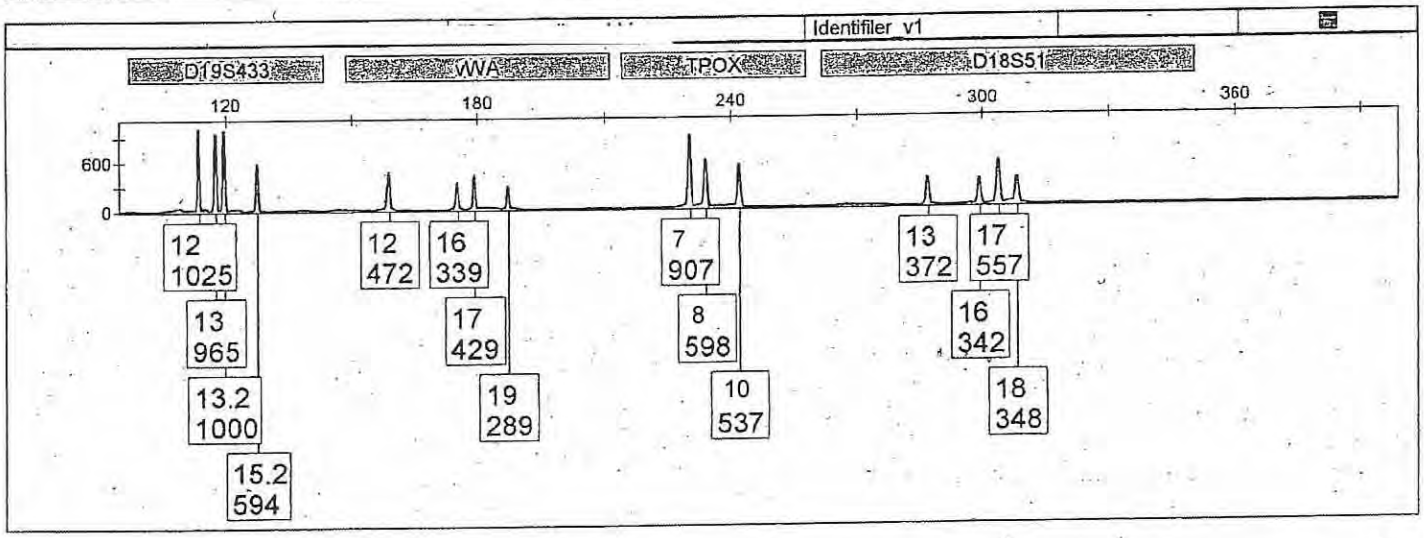
EVIDENCE SAMPLE (page 1 of 2)

14

Sample File	Sample Name	Panel	SQO	SQ
		Identifiler_v1		







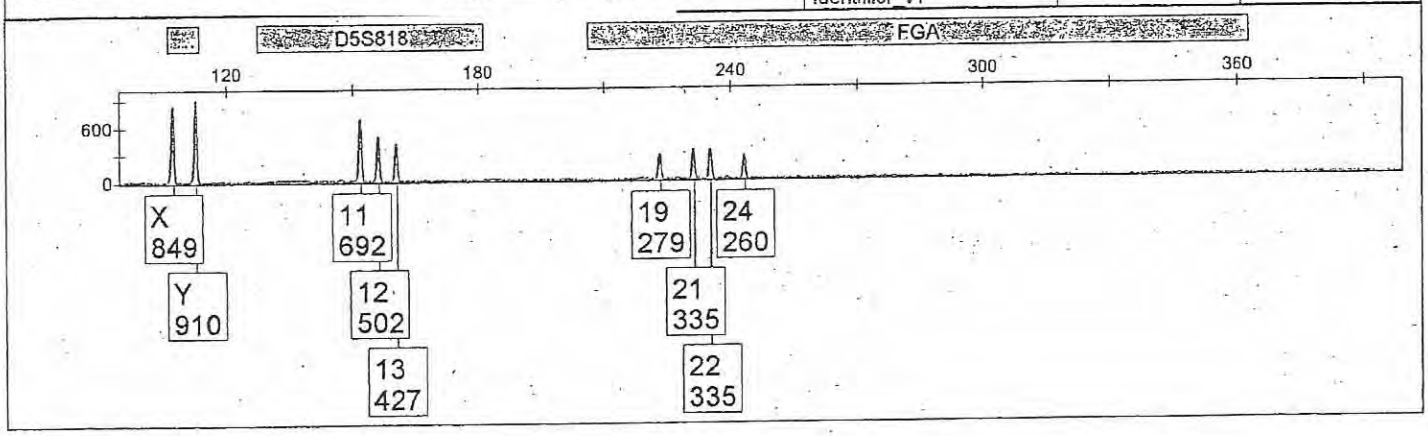
Applied Biosystems

GeneMapper ID v3.2.1

EVIDENCE SAMPLE (page 2 of 2)

15

Sample File	Sample Name	Panel	SQO	SQ
		Identifiler_v1		

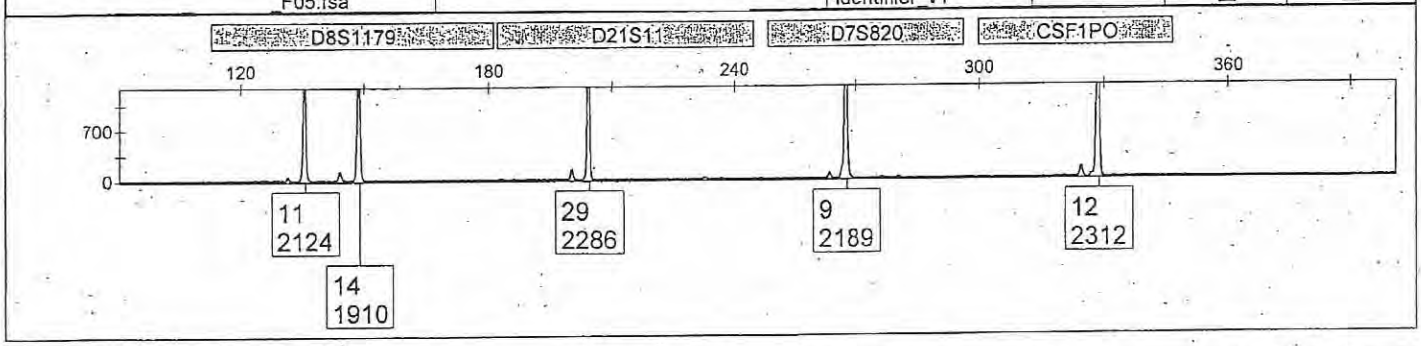


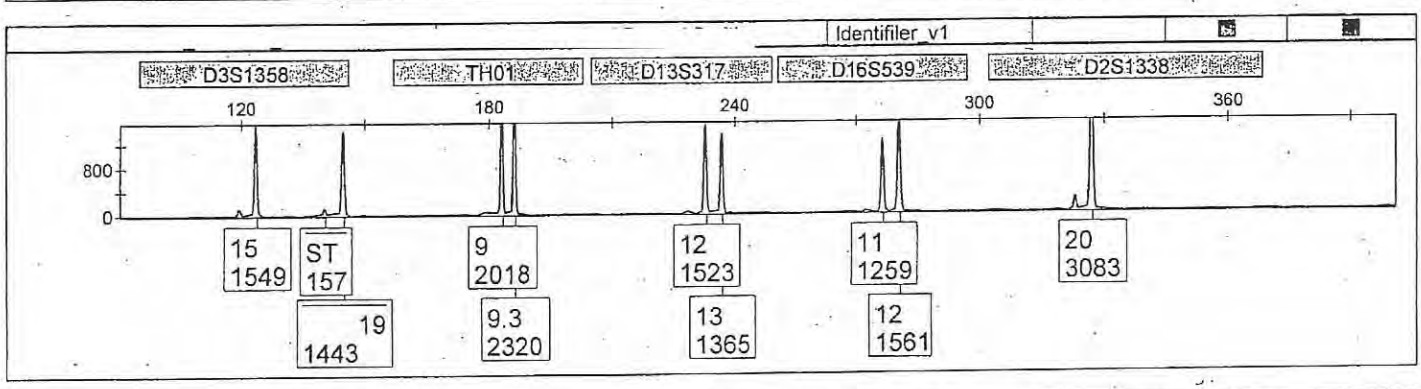
Applied Biosystems
GeneMapper ID v3.2

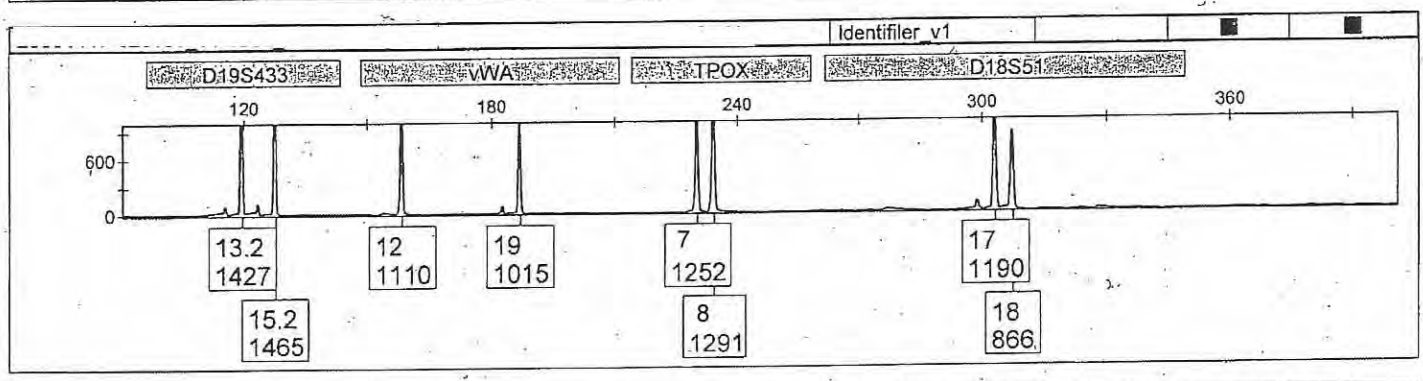
SUSPECT DNA REFERENCE SAMPLE

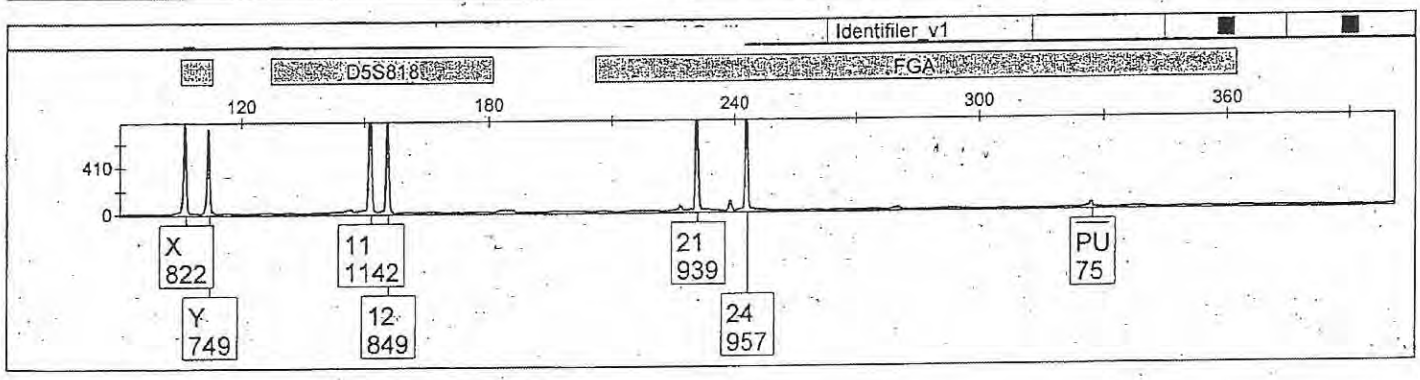
12

Sample File	Sample Name	Panel	SQO	OS	SQ
F05.fsa		Identifiler_v1			









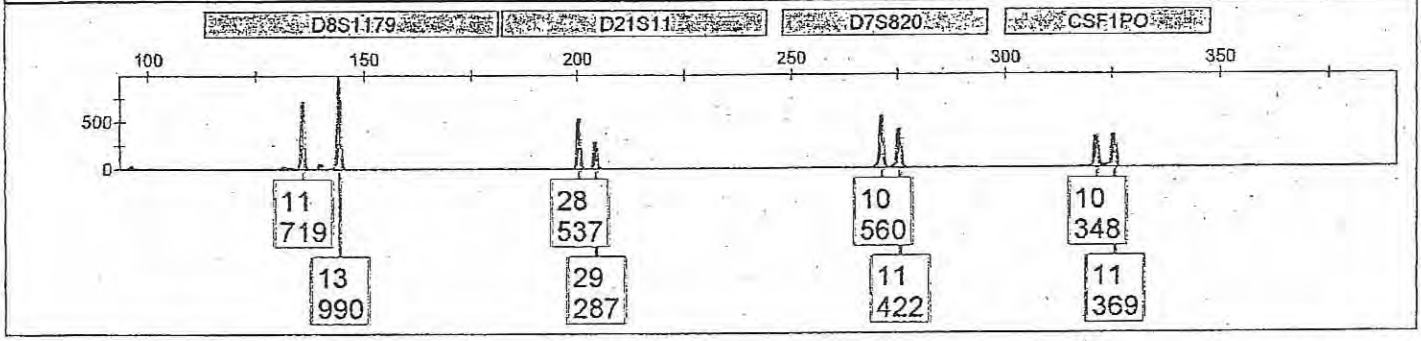
Printed by:

Applied Biosystems
GeneMapper ID v3.2.1

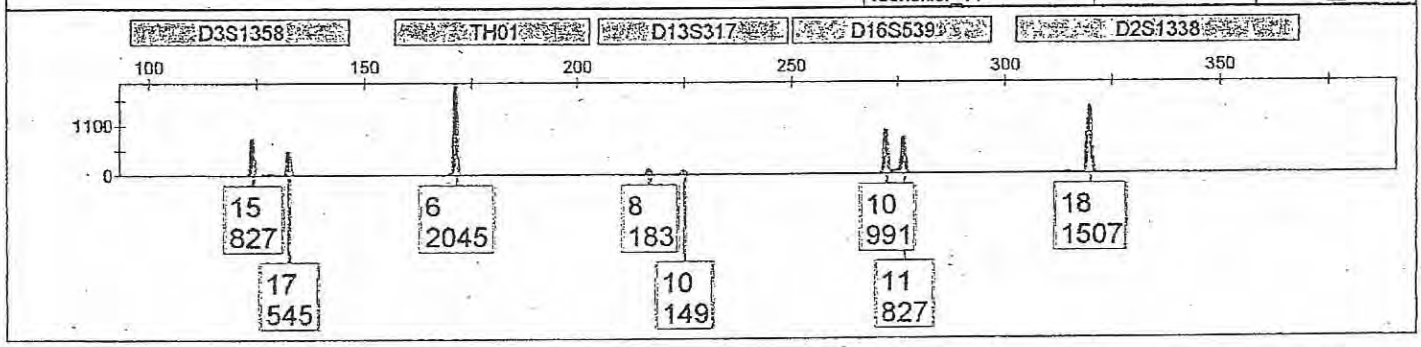
VICTIM DNA REFERENCE SAMPLE

21

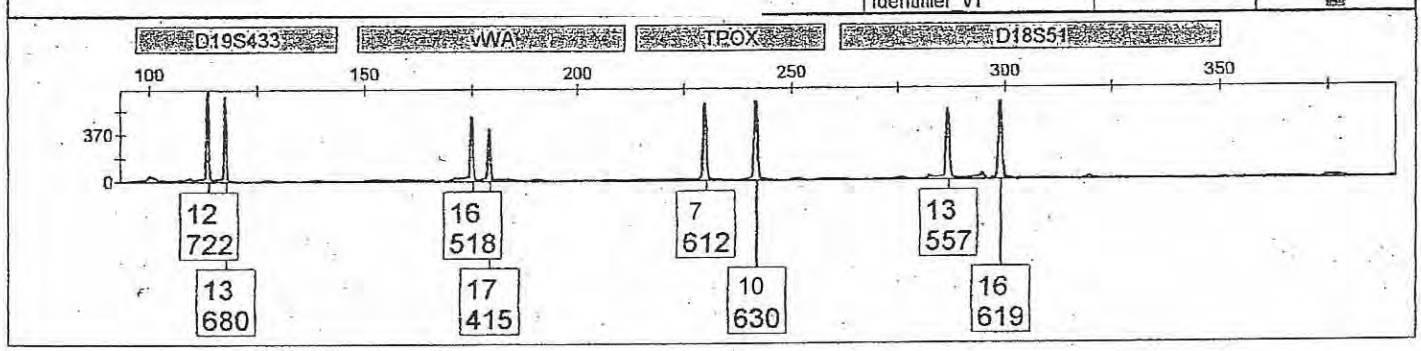
Sample File	Sample Name	Panel	SQO	SQ
		Identifiler_v1		



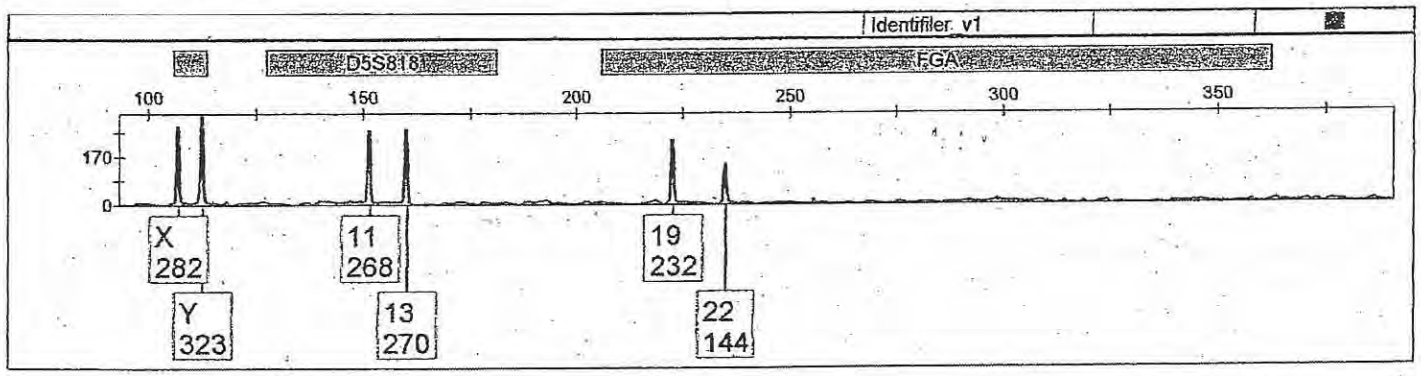
		Identifiler_v1		



		Identifiler_v1		



		Identifiler_v1		



ALLELE CALL TABLES

Item Number	14-15			
Item Description	Evidence Sample			
Markers	**Allele 1**	**Allele 2**	**Allele 3**	**Allele 4**
D8S1179	11	13	14	-
D21S11	28	29	-	-
D7S820	9	10	11	-
CSF1PO	10	11	-	-
D3S1358	15	17	19	-
TH01	6	9	9.3	-
D13S317	8	10	-	-
D16S539	10	11	12	-
D2S1338	18	20	-	-
D19S433	12	13	13.2	15.2
vWA	12	16	17	19
TPOX	7	8	10	-
D18S51	13	16	17	18
AMEL	X	Y	-	-
D5S818	11	12	13	-
FGA	19	21	22	24

Item Number	12				21			
Item Description	DNA from suspect				DNA from victim			
Markers	**Allele 1**	**Allele 2**	**Allele 3**	**Allele 4**	**Allele 1**	**Allele 2**	**Allele 3**	**Allele 4**
D8S1179	11	14	-	-	11	13	-	-
D21S11	29	-	-	-	28	29	-	-
D7S820	9	-	-	-	10	11	-	-
CSF1PO	12	-	-	-	10	11	-	-
D3S1358	15	19		-	15	17	-	-
TH01	9	9.3	-	-	6	-	-	-
D13S317	12	13	-	-	8	10	-	-
D16S539	11	12	-	-	10	11	-	-
D2S1338	20	-	-	-	18	-	-	-
D19S433	13.2	15.2	-	-	12	13	-	-
vWA	12	19	-	-	16	17	-	-
TPOX	7	8	-	-	7	10	-	-
D18S51	17	18	-	-	13	16	-	-
AMEL	X	Y	-	-	X	Y	-	-
D5S818	11	12	-	-	11	13	-	-
FGA	21	24		-	19	22	-	-