



## APPLICATION OF STR MARKERS IN FORENSICS

Anjali Somanathan<sup>1</sup>, Dr. Komal Mathur<sup>2</sup>

<sup>1,2</sup> Amity Institute of Biotechnology, Amity University, Sector 125, Noida, Uttar Pradesh (India)

### ABSTRACT

*In the modern day's criminal activities, the investigators usually encounter the evidence samples containing a trace biological material which even gets destroyed and gradually gets contaminated during the course of handling, transportation and subsequent analysis. This article explores the application of STR markers in DNA analysis. STRs have become popular DNA markers as they are easily amplified by polymerase chain reaction (PCR). The significant advantages of using STR markers over other markers have led to their use in forensics to identify victim and the perpetrator. STRs with higher power of differentiation are chosen for human identification in forensic cases on a regular basis.*

**Keywords:** DNA Profiling, Genotyping, Polymerase Chain Reaction, STR

### I INTRODUCTION

The ability to isolate, amplify and generate the DNA templates recovered from trace evidence samples has proved to be a major challenging area in the field of Forensic DNA Fingerprinting. In the modern day's criminal activities, the investigators usually encounter the evidence samples containing a trace biological material which even gets destroyed and gradually gets contaminated during the course of handling, transportation and subsequent analysis. However the reports of capability of recovery of amplifiable DNA samples from trace samples has come as a boon to the technology of forensic DNA profiling resulting in accurate individualization and matching.

The human genome is full of repeated DNA sequences. DNA regions with short repeat units (generally 2-6 bp in length) are called Short Tandem Repeats (STR). STRs are found surrounding the chromosomal centromere. STRs have proven to have several benefits that make them especially suitable for human identification.

STRs have become reliable DNA markers because they are easily amplified by polymerase chain reaction (PCR). An individual inherits one copy of an STR from each parent that may or may not have similar repeat sizes. The number of repeats in STR markers can be greatly variable among individuals, which make these STRs effective for human identification purposes. For human identification purposes, it is necessary to have DNA markers that exhibit the highest possible variation in order to discriminate between samples. It is often difficult to obtain PCR amplification products from forensic samples because either the DNA in the samples is degraded, or mixed, such as in a sexual assault case. The smaller size of STR alleles makes STR markers better for use in forensic applications, in which degraded DNA is common. PCR amplification of degraded DNA samples can be better off with smaller



target product sizes. Because of their smaller size, STR alleles can also be removed from other chromosomal locations more easily to ensure closely linked loci are not chosen. Closely linked loci do not associate with the predictable pattern of random distribution in the population, making statistical estimation difficult. STR alleles also have decreased mutation rates, which makes the data more stable and predictable. Due to these properties, STRs with higher power of discrimination are chosen for human identification in forensic cases on a regular basis. It is used to determine victim, perpetrator, missing persons, and others.

Beginning in 1996, the FBI Laboratory initiated a nationwide forensic science effort to establish core STR loci for incorporation within the national database called CODIS (Combined DNA Index System). The 13 CODIS loci are CSF1PO, FGA, TH01, TPOX, VWA, D8S1179, D3S1358, D16S539, D5S818, D7S820, D13S317, D18S51 and D21S11. These loci are nationally and internationally recognized as the standard for human identification. The DDC (DNA Diagnostics Center) Forensic Laboratory regularly uses the 13 CODIS loci and has additional loci for a vast and powerful STR testing battery if required.

## II REVIEW OF LITERATURE

### 2.1 DNA Profiling “Historical Background”

Forensic science has come a long way right from the use of blood groups in ruling out a suspect, to DNA fingerprinting in identifying a person. Forensic scientists used to employ various blood group systems like ABO, MNS system and Rh factor for an individual's identification but this technique only helped to exclude a suspect in forensic cases, reason being that the probability of matching of all blood groups occurring in a population is high throughout the world. The other markers used for forensic cases were protein and iso-enzyme markers. The disadvantage in using these conventional techniques is that protein markers have low efficiency. Band patterns resolved by iso-enzyme markers have the problem of low resolution, and fingerprints could not be obtained from much decomposed bodies and requires earlier recording or data.

It was in 1985 that professor Sir Alec Jeffreys from Leicester University, U.K. introduced the technique of DNA fingerprinting while studying myoglobin gene. He observed that nucleotides were repeated several times in a specific combination and termed such sequences as Variable Number of Tandem Repeats (VNTR). His study also showed that these VNTRs varied from an individual to individual and could be identified using restriction enzymes and the technique was known as Restriction Fragment Length Polymorphism (RFLP). This very case demonstrated the potential of DNA profiling and pointed towards its future as the most important forensic investigation tool to be developed in 20th Century.

DNA profiling has undergone three major stages of technological advancement. Loosely speaking these are *single-locus*, *multi-locus* and *STR stages*. The first method developed for forensic examination of samples pioneered by Dr. Jeffreys was termed as multi-locus testing. He observed that certain regions of DNA contained DNA sequences that were repeated over and over recurrently next to each other. He also discovered that the number of repeated sections



present in a sample could differ from individual to individual. These DNA repeat regions became known as “*mini satellites*” or *VNTRs*, which stands for *variable number of tandem repeats*. The method used by Dr. Jeffreys to examine the VNTRs was called Restriction Fragment Length Polymorphism (RFLP) as it involved the use of restriction enzymes to cut the regions of DNA surrounding the VNTRs. These fragments were electrophoresed on gel that separates them by size and then visualized using multi locus probes that hybridized many mini satellite loci at once.

This technique requires at least 500 mg of DNA, which, in trace biological samples, is difficult to retrieve. A newer form of analysis, Polymerase Chain Reaction (PCR), can be conducted with as few as, say, 100 white blood cells, found in an almost invisible speck of blood, compared to the thousands or even millions needed for RFLP. For PCR, smaller sample is required because it doubles and redoubles a single strand of DNA about 30 times, producing more than a billion copies. It utilizes “*micro satellites*” or *STRs* which stands for *short tandem repeats*. They are short sequences of 100 to 200 bp given by the repetition of 1-6 bp sequences.

### **2.1.1 SHORT TANDEM REPEATS (STRS)**

The human genome is full of contiguous repeated DNA sequences. These repeated sequences come in different sizes and are grouped according to the length of the core repeat units, the number of contiguous repeat units, and/or the overall length of the repeat region [1]. Long repeat units may have several hundred to several thousand bases in the core repeat. These regions are commonly referred to as satellite DNA and may be found around the chromosomal centromere. The core repeat unit for a medium length repeat, commonly referred to as a mini-satellite or a VNTR (variant number of tandem repeats), is in the range of roughly 10-100 bases in length [2, 3]. The forensic DNA marker D1S80 is a mini-satellite with a 16 bp repeat unit and contains alleles covering the range of 16-41 repeat units.

DNA regions with repeat units that are around 2-6 bp in length are called microsatellites, simple sequence repeats (SSRs), or short tandem repeats (STRs). STRs have become widely accepted DNA repeat markers because they are easily amplified by the polymerase chain reaction without the complication of differential amplification. This is due to the fact that both alleles from a heterozygous individual are alike in size since the repeat size is small. The number of repeats in STR markers can be greatly variable among individuals.

Microsatellites account for roughly 3% of the total human genome. STR markers are scattered throughout the genome and appear on average every 10000 nucleotides [2, 4].

### **2.1.2 ISOLATION AND TYPES OF STR MARKERS**

For the analysis of STR markers, the invariant flanking regions surrounding the repeats must be determined. Once the flanking sequences are established then PCR primers can be designed and the repeat region amplified for analysis. New STR markers are usually identified in one or of two ways:



(1) Seeking DNA sequence databases such as GenBank for regions with more than six or so adjoining repeat units [1, 4, 5]; or

(2) Performing molecular biology isolation methods [4, 5].

STR repeat sequences are named by the length of the repeat unit. Dinucleotide repeats have two nucleotides repeated next to each other continuously. Trinucleotides have three, tetranucleotides have four, pentanucleotides have five, and hexanucleotides contain six repeat units in the core repeat. Tetranucleotide repeats have become the most accepted STR markers for human identification.

STRs are often divided into several categories based on the repeat pattern:-

1. **Simple repeats** contain units of identical length and sequence
2. **Compound repeats** comprise two or more contiguous simple repeats
3. **Complex repeats** may contain numerous repeat blocks of variable unit length as well as variable intervening sequences [6].
4. **Complex hypervariable repeats** also exist with several non-consensus alleles that differ in both size and sequence and are difficult to genotype reproducibly [7, 8]. This category of STR markers is not as widely used in forensic DNA typing due to problems with allele nomenclature and measurement variability between various laboratories [8].

Microvariants are alleles that have incomplete repeat units. One example of a microvariant is allele 9.3 at the TH01 locus which consists of nine tetranucleotide repeats and one incomplete repeat of three nucleotides as the seventh repeat is lacking a single adenine out of the normal AATG repeat unit.

Apparently there are 4, 16, 64, 256, 1024, 4096 likely motifs for mono-, di-, tri-, tetra-, penta-, and hexanucleotide repeats, respectively. However, as microsatellites are tandemly repeated, some motifs are exactly equivalent to others. Two rules can be used to identify if motif A is equivalent to motif B. Motif A is treated equivalent to motif B when (1) motif A is inversely complementary to motif B, or (2) motif A is distinct from motif B or the inversely complementary sequence of motif B by frameshift. For example, (GAAA)<sub>n</sub> is equivalent to (AGAA)<sub>n</sub> or (AAGA)<sub>n</sub> to (AAAG)<sub>n</sub>, or (TTTC)<sub>n</sub> to (TTCT)<sub>n</sub> or (TCTT)<sub>n</sub> or to (CTTT)<sub>n</sub>.

As a result of this equivalence in repeat motif structure there are only 2, 4, 10, 33, 102 and 350 likely motifs for mono-, di-, tri-, tetra-, penta-, and hexanucleotide repeats, respectively.

**Table 1- Types of repeats (based on number)**

Repeats	Repeat Number	Sequence
Mononucleotide repeats	2	A C
Dinucleotide repeats	4	AC AG AT CG
Trinucleotide repeats	10	AAC AAG AAT ACC ACT AGC AGG ATC



		ACG
Tetranucleotide repeats	33	AAAC AAAG AAAT AACC AACG AACT AAGC AAGG AAGT AATC AATG AATT ACAG AGAT AGCC AGCG AGCT AGGC AGGG ATCC ATCG ATGC ACAT ACCC ACCG ACCT ACGC ACGG ACGT ACTC ACTG CCCG CCGG

*AGAT or GATA motif is the most common for STR loci used by forensic scientists. [9]*

### 2.1.3 Desirable Characteristics of STRS Used In Forensic DNA Typing

For human identification purposes it is necessary to have DNA markers that present the highest possible variation or a number of less polymorphic markers that can be assembled to acquire the ability to discriminate between samples. Forensic specimens are usually challenging to PCR amplify because the DNA in the samples may be extremely degraded. Mixtures are prevalent in some forensic examples, such as those recovered from sexual assault cases consisting of biological sample from both the perpetrator and victim. Relatively larger size of mini-satellite VNTR alleles (~400-1000 bp) than STR alleles (~100-400 bp) and allelic drop out of mini-satellite markers caused by allelic dropout of larger alleles in mini-satellite markers produced by preferential amplification of the smaller allele is also a fundamental problem with mini-satellites.

Among the various types of STRs, tetranucleotide repeats have become more popular than di- or trinucleotides. A biological phenomenon known as 'stutter' results when STR alleles are PCR amplified. Stutter products are amplicons that are often one or more repeat units less in size than the true allele and arise during PCR because of strand slippage [10]. With di- and trinucleotides, the stutter percentage can be much higher (30% or more) making it difficult to interpret sample mixtures.

Advantages of using tetranucleotide STR loci in forensic DNA typing over VNTR mini-satellites or di- and trinucleotide repeat STR:

- A narrow allele size range that permits multiplexing;
- A narrow allele size range that decreases allelic dropout from preferential amplification of smaller alleles;
- The capacity of generating small PCR product sizes that assist recovery of information from degraded DNA specimens; and
- Decreased stutter product formation compared to dinucleotide repeats that assist the interpretation of sample mixtures.

The selection basis for candidate STR loci in human identification applications involves the following characteristics [11]:

- High discriminating power, generally >0.9, with observed heterozygosity >70%;



- Separate chromosomal locations to confirm that closely linked loci are not chosen;
- Robustness and reproducibility of results when multiplexed with other markers;
- Low mutation rate;
- Low stutter characteristics;
- Predicted length of alleles that fall in the range of 90-500 bp with smaller sizes better suited for interpretation of degraded DNA samples.

STR markers used in forensic DNA typing are generally chosen from separate chromosomes to avoid problems with linkage between the markers.

### 2.1.4 Allelic Ladders

An allelic ladder is an artificial mixture of the common alleles existing in the human population for particular STR marker [8]. Allelic ladders have proved to be important for accurate genotype determination. These allelic ladders provide a standard for measuring each STR locus.

Allelic ladders are composed by combining genomic DNA or locus-specific PCR products from several individuals in a population, which possess alleles that are representative of the variation for the given STR marker [8, 12]. The samples are then co-amplified to create an artificial sample containing the common alleles for the STR marker.

### 2.2 DNA Extraction

DNA extraction methods have been generated to separate proteins and other cellular materials from DNA molecules. There are three primary techniques for DNA extraction used in today's forensic DNA laboratory: organic extraction, Chelex extraction, and FTA paper [11]. The specific extraction or DNA isolation procedure varies depending on type of biological evidence being analyzed. For example, whole blood must be treated separately from a bloodstain or a bone fragment. Organic extraction, often referred to as phenol chloroform extraction, has been in use for the longest stretch of time and may be used for cases where either RFLP or PCR typing is executed. High molecular weight DNA, which is necessary for RFLP methods, may be obtained most adequately with organic extraction.

All samples must be delicately handled regardless of the DNA extraction method to block sample-to-sample contamination or addition of extraneous DNA. The extraction process is likely where the DNA sample is more susceptible to contamination in the laboratory than at any other period during the forensic DNA analysis process. Extracted DNA is generally stored at -20°C, or even -80°C for long-term storage, to inhibit nuclease activity. Nucleases are enzymes found in cells that degrade DNA and allow recycling of the nucleotide components. Nucleases need magnesium to work properly therefore one of the measures to prevent them from digesting DNA in blood is the utilization of purple-topped tubes consisting of a blood preservative known as EDTA. The EDTA chelates all of the free magnesium and therefore prevents the nucleases from completely destroying the DNA in the collected blood sample.





### 2.3 Quantification of Nucleic Acid

Quantification of nucleic acids is commonly used to determine the concentrations of DNA or RNA present in a mixture, as following reactions or protocols using a nucleic acid sample usually require particular amounts for optimum performance. There exist several techniques to provide the concentration of a solution of nucleic acids, in addition to agarose gel electrophoresis and spectrophotometric quantification.

#### Quality and Quantity Assessment of DNA Using Agarose Gel Electrophoresis:

Agarose gel electrophoresis is a basic way to separate DNA fragments by their sizes and visualize them [13, 14].

The method of electrophoresis is based on the fact that DNA is negatively charged at neutral pH because of its phosphate backbone. Therefore, when an electrical potential is established on the DNA it will move towards the positive pole. The rate at which the DNA will move towards the positive pole is slowed by preparing the DNA move through an agarose gel. This is a buffer solution (which maintains the adequate pH and salt concentration) with 0.75% to 2.0% agarose added. The agarose forms a porous lattice in the buffer and the DNA must slip through the holes in the lattice to move toward the positive pole. This slows down the molecule. Larger molecules will be slowed down more than smaller molecules, since the smaller molecules can fit through the holes easily [15]. Thus, a mixture of large and small fragments of DNA that has been run through an agarose gel will be isolated by size. Extracted DNA samples are allowed to run on an 1% agarose gel in 1X- TAE buffer. EtBr (Ethidium Bromide) punched in the gel intercalates between the double stranded DNA sample and fluorescence under UV light depending on the quantity of extracted DNA sample.

#### Components

**Agarose:** Agarose is isolated from agar, a gelatinous substance obtained from algae. Different purities of agarose are available commercially as are agaroses with various melting properties. High purity low melt agarose is generally used if the DNA is to be extracted from the gel [16].

**Buffers:** There are many buffers used for agarose electrophoresis. The most commonly used being: Tris acetate EDTA (TAE), Tris/Borate/EDTA (TBE) and Sodium borate (SB). TAE has the lowest buffering capacity but gives the best resolution for larger DNA. This means a lower voltage and more time to run [15].

**Ethidium Bromide (Etbr) And Dyes:** The universal dye used to make DNA or RNA bands visible for agarose gel electrophoresis is Ethidium bromide, often abbreviated as EtBr [10]. It fluoresces under UV light when intercalated into DNA (or RNA). By running DNA through an EtBr-treated gel and view it with UV light, distinct bands of DNA become visible. EtBr is a common carcinogen, however safer alternatives are available.

Loading buffers are added with the DNA to visualize and sediment it in the gel well. Negatively charged indicators track the location of the DNA. Xylene cyanol and Bromophenol blue are generally used. They run at about 5000 bp and 300 bp respectively, but the exact position varies with percentage of the gel. Other less commonly used progress markers are Cresol Red and Orange G which run at about 125 bp and 50 bp [17].

**Analysis:** After electrophoresis, the gel is illuminated with an ultraviolet lamp to view the DNA bands. The ethidium bromide fluoresces reddish-orange in the presence of DNA. The DNA band can be cut out of the gel, and can then be dissolved to retain the purified DNA. The gel can then be photographed generally with a digital or polaroid camera [14, 18]. Even though the stained nucleic acid fluoresces reddish-orange, images are shown in black and white.

## 2.4 DNA Amplification & Polymerase Chain Reaction

Forensic science and DNA typing laboratories have largely advanced from the discovery of a technique known as polymerase chain reaction or PCR. The invention of PCR has brought revolution in molecular biology which has enabled to make million of copies of a specific sequence of DNA in a matter of only a few hours. The PCR DNA amplification technology is well suited to analysis of DNA samples because it is sensitive, rapid, and not limited by the quality of the DNA as are the restriction fragment length polymorphism (RFLP) methods.

### 2.4.1 PCR PROCESS

PCR is an enzymatic method in which a specific region of DNA is replicated over and over again to yield many copies of a particular sequence [19, 20]. This molecular 'xeroxing' process involves heating and cooling samples in a precise thermal cycling pattern over ~ 30 cycles. Theoretically after 28 cycles approximately a billion copies of the target region on the DNA template have been generated. The PCR product, referred to as an 'amplicon', is then in sufficient quantity that it can be easily measured. PCR is commonly performed with a sample volume in the range of 5-100µL.

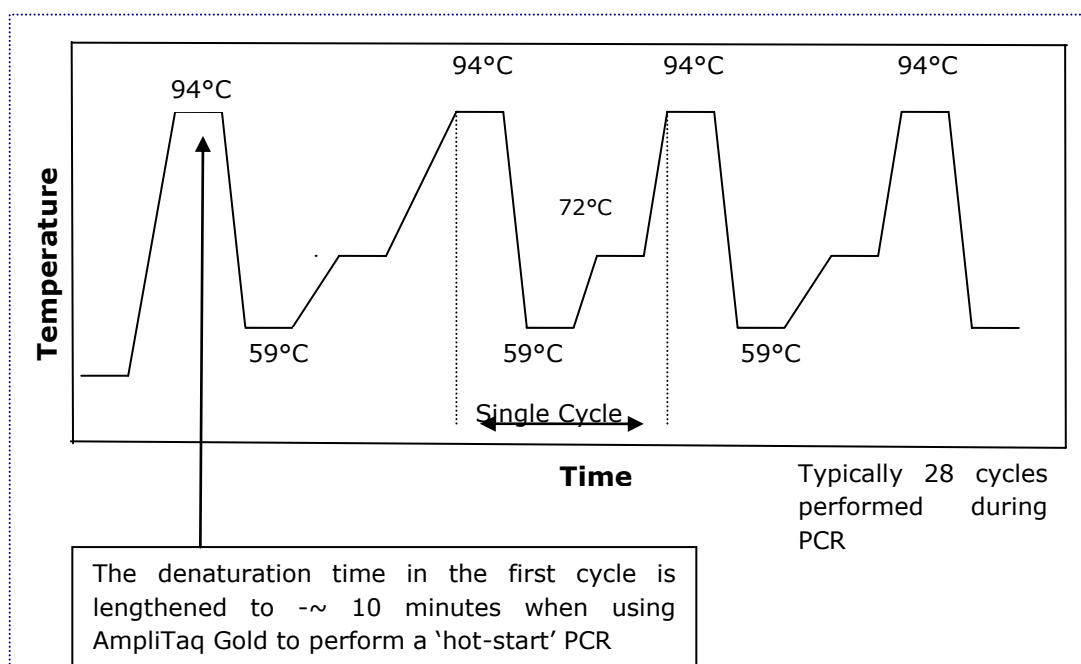


Fig. 1- Thermal cycling temperature profile of PCR





Thermal cycling typically involves three different temperatures that are repeated over and over again 25-35 times. At 94°C, the DNA strands separate, or 'denature'. At 60°C, primers bind or 'anneal' to the DNA template and target the area to be amplified. At 72°C, the DNA polymerase extends the primers by copying the target area using the deoxyribonucleotide triphosphate building blocks. The entire PCR process is about 3 hours with each cycle taking ~5 minutes on traditional thermal cyclers: 1 minute each at 94°C, 60°C, and 72°C and about 2 minutes ramping between the three temperatures.

### 2.4.2 PCR COMPONENTS

A PCR reaction is prepared by mixing several individual components and then adding de-ionized water to achieve the required volume and concentration of each of the components. Commercial kits with pre-mixed components may also be used for PCR. These kits have largely simplified the use of PCR in forensic DNA laboratories. **PCR contains four essential components:**

- **A Thermostable DNA Polymerase:**

(Catalyze Template Dependent Synthesis of DNA)

A wide choice of enzymes is now available that vary in the fidelity, efficiency and ability to synthesize large DNA products. For routine PCRs *Taq* polymerase remains the enzyme of choice. The **Taq polymerase** is a thermostable DNA polymerase named after the thermophilic bacterium *Thermos aquaticus* from which it was first obtained. It's normally abbreviated to "TaqPol" or simply "Taq". The enzyme, similar to other DNA polymerases, catalyzes template-directed synthesis of DNA for nucleotide triphosphates. A primer having a free 3' hydroxyl is needed to initiate synthesis and magnesium ion is necessary. A modified form of Taq DNA polymerase has been formed that requires thermal activation and thus enables a closed-tube hot start PCR. This enzyme, AmpliTaq Gold has immensely benefited the specificity of PCR amplifications.

- **Amplitaq Gold DNA Polymerase**

AmpliTaq Gold™ DNA polymerase is a chemically modified enzyme that is rendered inactive until heated [21]. An extended pre-incubation of 95°C, normally for 10 or 11 minutes, is used to activate the AmpliTaq Gold. The chemical modification contains a derivitization of the epsilon-amino groups of the lysine residues [22]. At a pH below 7 the chemical modification moieties get detached and the activity of the polymerase is restored. The Tris buffer in the PCR reaction is pH sensitive with temperature variation, and higher temperatures cause the solution pH to go down by approximately by 0.02 pH units with every 1°C [23]. It is important to note that AmpliTaq Gold is not compatible with the pH 9 buffers used for regular AmpliTaq DNA polymerases [24]. This fact is because the pH does not get low enough to remove the chemical modification on Taq Gold and thus the enzyme remains inactive. Tris buffer with a pH 8.0 or 8.3 at 25°C work the best with Taq Gold.

- **A Pair of Synthetic Oligonucleotides:**

(Priming the DNA Synthesis)



Careful design of primers is required to get the desired products in high yield, to decrease amplification of unwanted sequences and to aid subsequent manipulation of the amplified product. PCR yield is affected by the annealing characteristics of the primers. For the PCR to work effectively, the two primers must be specific to the target region, obtain similar annealing temperatures, not interact considerably with each other or themselves to build 'primer dimers', and be structurally compatible.

- **Deoxynucleoside Triphosphates (dNTPS):**

Standard PCRs have equimolar amounts of dATP, dTTP, dCTP, and dGTP. Concentrations of 200-250  $\mu\text{M}$  of each dNTP are included for Taq polymerase in reactions containing 1.5 mM  $\text{MgCl}_2$ . In a 50 $\mu\text{l}$  reaction, these amounts should allow synthesis of ~6-6.5  $\mu\text{g}$  of DNA, which should be adequate even for multiplex reactions in which eight or more primer pairs are used simultaneously. High concentrations of dNTPs (>4 mM) are inhibitory, perhaps because of sequestering of  $\text{Mg}^{2+}$ . However, a sufficient amount of amplified product can be produced with dNTP concentrations as low as 20 $\mu\text{M}$ -0.5-1.0 pmole of an amplified fragment ~1kb in length. To avoid problems, stocks of dNTPs (100-200 mM) should be discarded after the second cycle of freezing/thawing.

- **Template DNA:**

Template DNA having target sequences can be added to PCR is single or double stranded form. Closed circular DNA templates are amplified slightly less effectively than linear DNAs. Although the size of the temple DNA is not critical, amplification of sequences fixed in high-molecular-weight DNA (> 10 kb) can be improved by digesting the template with a restriction enzyme that do not cleave within the target sequence. When working at its best, PCR requires only a single copy of a target sequence as template. More typically, however, several thousand copies of the target DNA are seeded into the reaction. In the case of Human genomic DNA, up to 1.0  $\mu\text{g}$  to 2.0 $\mu\text{g}$  of DNA is utilized per reaction.

### 2.4.3 PHASES OF POLYMERASE CHAIN REACTION

- **PHASE 1: Hot Start**

Low temperature mispriming can be avoided by starting PCR at an elevated temperature, a process usually referred to as 'hot start' PCR. Hot start PCR may be done by introducing a critical reaction component such as polymerase, after the temperature of the sample has been elevated above the desired annealing temperature (example: 60°C). This decreases the possibilities of mispriming and misextension events by not having the polymerase present during reaction setup.

- **PHASE 2: Denaturation**

Double stranded DNA (ds DNA) templates denature at a temperature that is determined in part by their G+C content. The greater the proportion of G+C, the greater the temperature required to separate the strands of template DNA. The longer the DNA molecules, the longer the time required at the chosen denaturation temperature to separate the two strands entirely. If the temperature for denaturation is too small or if the time is too brief only AT-rich regions of the template DNA will be denatured. In PCRs activated by Taq DNA polymerase, denaturation is carried out at 94-95°C, which is the maximum temperature that the enzyme can withstand for 30 or more cycles



without sustaining excessive damage. The denaturation is carried out for about 45 seconds at 94-95°C for conventional amplification of linear DNA templates.

- **PHASE 3: Annealing**

The temperature used for the annealing of primers to template DNA is critical. If the annealing temperature is too steep, the oligonucleotide primers anneal poorly, if not all, to the template and the yield of amplified DNA is very small. If the annealing temperature is too small, non-specific annealing of primers may occur resulting in the amplification of non-specific segments of DNA. Annealing is usually carried out 3-5°C lower than the calculated melting temperature at which the oligonucleotide primers detach from their templates.

- **PHASE 4: Extension**

Extension of oligonucleotide primers is executed at or near the optimal temperature for DNA synthesis catalyzed by the thermostable polymerase, in the case of Taq DNA polymerase which is 72-78°C. In the first two cycles, extension from one primer precedes beyond the sequence complementary to the binding site the other primer. In the next cycles, the primary molecules are produced whose length is equal to the segment of the DNA delimited by the binding sites of the primers. From the third cycle onwards the segment of the DNA is amplified geometrically, whereas longer amplification products accumulate arithmetically. The rate of polymerization of Taq polymerase is ~2000 nucleotides per minute at the optimal temperature (72-78°C).

### **NUMBER OF CYCLES**

The number of cycles needed for amplification depends on the number of copies of template DNA present at the beginning of the reaction and the capability of primer extension and amplification. Once established in the geometric phase, the reaction proceeds until one of the component becomes limiting. At least 25 cycles are required to achieve acceptable levels of amplification of single copy target sequence in mammalian DNA templates.

### **PCR INHIBITORS AND DNA DEGRADATION**

When extracting biological materials for the aim of forensic DNA typing, it is important to try and avoid further degradation of the DNA template and to remove inhibitors of the polymerase chain reaction (PCR) where possible. The presence of inhibitors or degraded DNA can exhibit themselves by complete PCR amplification failure of a reduced sensitivity of detection often for the larger PCR products.

Two PCR inhibitors usually found in forensic cases are hemoglobin and indigo dyes from denim. Melanin recovered in hair samples can be a source of PCR inhibition when trying to amplify mitochondrial DNA. This inhibitor likely bind in the active site of the Taq DNA polymerase and prevent its proper functioning during PCR amplification.

DNA degrades through numerous mechanisms including both enzymatic and chemical processes [25]. Once a cell (or organism) dies its DNA molecules face cellular nucleases followed by bacterial, fungal, and insect onslaughts depending on the environmental conditions [26]. In addition, hydrolytic cleavage and oxidative base damage can restrict successful recovery and amplification of DNA. The main target of hydrolytic cleavage is the glycosidic base sugar bond. Breakage of this bond leads to nucleobase loss and then a single stranded 'nick' at the basic site. If adequate number of DNA molecules in the biological sample dissociate in a region where primers anneal or



between the forward and reverse primers, then PCR amplification efficacy will be decreased or the target region may fail to be amplified at all. Thus heat and humidity which pace up hydrolytic cleavage are enemies of intact DNA molecules. Furthermore, UV irradiation can lead to cross-linking of adjacent thymine nucleotides on the DNA molecule, which will prevent passage of the polymerase during PCR.

#### **2.4.4 MULTIPLEX PCR**

The polymerase chain reaction allows more than one region to be copied simultaneously by simply adding more the one primer set to the reaction mixture [2]. The simultaneous amplification of two or more regions of DNA is commonly known as multiplexing. For a multiplex reaction to work accurately the primer pairs need to be compatible. In other words, the primer annealing temperatures should be similar and excessive regions of complementarities should be avoided to prevent the production of primer-dimers that will cause the primers to bind to one another instead of the template DNA.

In fact, multiplex PCR optimization is more of a challenge than single-plex reactions because so many primer-annealing events must occur at the same time without interfering with each other. The variables that are examined when trying to obtain best results for a multiplex PCR amplification include many of the reagents as well as the thermal cycling temperature profile. Extension times during thermal cycling are often increased for multiplex reactions in order to give the polymerase time to completely copy all of the DNA targets. In addition, the primers must produce robust amplifications with acceptable peak height balance between loci as well as specific amplification with o non-specific products that might interfere with suitable interpretation of a sample's DNA profile. Finally, primers should produce a maximal non-template-dependent '+A' addition to all PCR products.

#### **Advantages of PCR with Forensic Specimens**

The advantages of PCR amplification for biological evidence include the following:

- DNA degraded to fragments only a hundred base pairs in length can serve as efficient for amplification.
- Very little amounts of DNA template may be used from as small as a single cell.
- Contaminant DNA, such as fungal and bacterial sources, will not amplify as human-specific primers are used.
- Larger numbers of copied of specific DNA sequences can be amplified at the same time with multiplex PCR reactions.
- Commercial kits are now available for easy PCR reaction setup and amplification.

#### **2.5 Genotyping**

Multiplex PCR amplification of a subset or all of these STR markers is possible with different commercial STR kits using spectrally resolvable fluorescent dyes.

**Table 2- Commercial STR kits**

Dye	6-FAM	VIC	NED	PET



Loci	CSF1P0	D2S1338	D18S51	D5S818
	D7S820	D3S1358	D19S433	FGA
	D8S1179	D13S317	TPOX	Amelogenin
	D21S11	D16S539	vWA	
		TH01		

The availability of commercial STR kits has largely clarified the usage of STRs in recent years and helped the development of large and efficient DNA databases. A report by the National Commission about the Future of DNA Evidence resolves that STR typing will likely be the dominant means of forensic DNA analysis for the next 5–10 years due to the need for flexibility in national and international DNA databases. STR markers offer a large number of advantages over previously used methods for DNA typing including the ability to gather results from degraded DNA samples and very small amounts of DNA. The process is fairly rapid and results may regularly be obtained in less than one working day.

PCR products are separated through size and dye color using electrophoresis followed by laser-induced fluorescence with multi-wavelength detection. An internal standard contains DNA fragments of known size and labeled with a different dye color, is usually co-electrophoresed with each sample to calibrate sizes from run to run. The collected data in the form of multicolored electropherograms are evaluated by software that automatically figures STR allele sizes based on a standard curve obtained from the internal size standard. STR genotyping is performed by comparing the allele sizes in each sample to the sizes of alleles available in an allelic ladder, which contains common alleles that have been already sequenced. On a capillary electrophoresis (CE) system, the allelic ladder is run along with the internal size standard in one injection, and allelic samples with the same internal size standard are run in consequent injections on the capillary in a sequential fashion.

In order to precisely genotype STR markers using multicolor fluorescence detection, a separation and detection technique must show the following characteristics:

- Methods for reliable sizing over a 75–500 bp size range.
- Effective color separations of different dye sets used to prevent bleed through between four or five different colors.
- High run-to-run precision between processed samples to allow comparison of allelic ladders to sequentially processed STR samples.
- Resolution of at least 1 bp to roughly 350 bp to permit reliable detection of micro variant alleles.

## The Need for DNA Separations

A polymerase chain reaction (PCR) reaction in which short tandem repeat (STR) alleles are amplified produces a mixture of DNA molecules that present a challenging separation problem. A multiplex PCR can produce 20 or more DNA fragments that must be resolved from one another.



In order to distinguish the various molecules from one another, a separation step is needed to pull the different sized fragments apart. The separation is typically performed by a process known as electrophoresis and is conducted in a slab-gel or capillary environment.

### **2.5.1 ELECTROPHORESIS**

PCR products from short tandem repeat DNA must be separated in a fashion that allows each allele to be different from other alleles. Heterozygous alleles are resolved in this manner with a sized based separation method known as **electrophoresis**. The separation medium may be in the form of a slab gel or a capillary.

### **2.5.2 CAPILLARY ELECTROPHORESIS**

Capillary electrophoresis (CE) is a new addition to the electrophoresis family. The first CE separations of DNA were performed just over a decade ago in the late 1980s [27]. There are a number of advantages to analyzing DNA in a capillary format.

#### **Advantages of CE over Slab Gels**

- Firstly, the injection, separation, and detection steps can be fully automated permitting multiple samples to be run unattended.
- Only minute quantities of sample are consumed in the injection process and samples can be easily retested if needed. This is an important advantage for precious forensic specimens that often cannot be easily replaced.
- Separation in capillaries may be conducted in minutes rather than hours due to higher voltages that are allowed with improved heat dissipation from capillaries.
- Quantitative information is readily available in an electronic format following the completion of a run.
- There is no fear of cross-contamination from samples leaking over from adjacent cells with CE as in the case of slab gel electrophoresis.

**Components of CE:** The primary CE instruments include: A narrow capillary, two buffer vials, two electrodes connected to a high-voltage power supply, fluorescence detector, laser excitation source, autosampler to hold the sample tubes and a computer to control the sample injection and detection.

The capillary is a narrow glass tube almost 50 cm long and 50  $\mu\text{m}$  in diameter. It is filled with a viscous polymer solution that acts much like a gel creating a sieving environment for DNA molecules. Samples are placed into a tray and injected onto the capillary by enabling a voltage to each and every sample sequentially.

A high voltage (e.g., 15000 volts) is applied across the capillary after the injection to separate the DNA fragments in a matter of minutes. Fluorescent dye-labeled products are evaluated as they pass by the detection window and are excited by a laser beam. Computerized data acquisition facilitates rapid analysis and digital storage of separation





results. Electronic fields are on the order of 10-to-100 times stronger with CE than with gels which results in faster run times for CE.

Detection of sample is performed by the CE instrument through measuring the time span from sample injection to sample detection with a laser kept near the end of the capillary. Laser light is shined on to the capillary at a fixed position where a window has been burned in the coating of the capillary. DNA fragments are illuminated as they pass by this window in the capillary. As with gels, the smaller molecules will arrive at the detection point first followed by the larger molecules. Data from CE separations are plotted as a function of the relative fluorescence intensity detected from fluorescence emission of dyes passing through the detector. The fluorescent emission signals from dyes attached to the DNA molecules can then be used to observe and quantify the DNA molecules passing the detector.

### **2.5.3 SAMPLE PREPARATION & INJECTION**

Samples for CE separation are usually made by diluting a small portion of the PCR product into de-ionized formamide. Another important advantage for CE in the context of forensic analysis is that only a small portion of the actual sample is examined each time.

Since formamide is a strong denaturant, it is commonly used in the preparation of single-stranded DNA samples for CE. Simply placing a sample in formamide is sufficient to denature it [28]. However, rapid heating to 95°C and snap-cooling on ice is often performed to ensure that the denaturation process has occurred. Use of high-quality formamide with a low conductivity is necessary.

### **2.5.4 ELECTROPHORETIC SEPARATIONS**

Electrophoresis is a relative rather than an absolute measurement technique. The position of a DNA band on a gel has no meaning without reference to a size standard containing material with known DNA fragment sizes. Thus, samples are run on a gel side-by-side with molecular weight markers. A visual comparison can then be made to estimate the fragment size of the unknown sample based on which band it comes closest to since the samples were subjected to identical electrophoretic conditions.

#### **Sample preparation**

Besides the width of the sample injection zone, there are several other components that influence DNA separations within CE systems: the polymer used for enabling the separation, the capillary, the electrophoresis buffer, and the field strength. STR allelic ladders are useful tools for monitoring system resolution [29].

#### **The polymer separation matrix**

There are several types of sieving media utilized in electrophoretic separations, depending on the physical characteristics of the media.

- Chemical gels such as the common polyacrylamide gels used in denaturing slab gel electrophoresis are solid cross-linked materials whose porous structure is linked together by strong covalent bonds.



- Agarose produces physical gels. The material's shape is the result of weaker intermolecular forces produced via entanglement of the diverse strands of different agarose molecules.

Entangled polymers are the third category of sieving media. Similar to physical gels, these materials are also characterized by intermolecular interactions. However, those substances are not true gels, as they cannot hold their shape unless placed in some container such as a capillary [30]. Entangled polymers are characterized by a rapid increase in viscosity as the polymer concentration reaches a definite threshold value.

The idea of using entangled polymer solutions to separate biopolymers is not new, as it was proposed years ago by Bode. However, it only became popular in combination with CE, because the very efficient anti-convective and heat dissipation properties of thin capillaries allow separation in fluids without loss of resolution. Grossman and Soane proved that by using a dilute, low-viscosity polymer solution as the separation medium, high-resolution separations of DNA mixtures could be achieved. The ideal polymer should have at least the same separation properties as classical gels, combined with a low viscosity that would permit easy replacement [31]. These conditions have been achieved with the performance optimized polymers, such as POP-4 and POP-6, from Applied Biosystems. POP-4 is commonly used for DNA fragment assay along with STR typing while the POP-6 polymer, which is the same polydimethylacrylamide polymer present at a higher concentration, is able to achieve higher resolution to meet the single-base resolution needs of DNA sequencing [32].

### The Buffer

The buffer which is used to dissolve the polymer in CE systems is important as it stabilizes and solubilizes the DNA, supplies charge carriers for the electrophoretic current, and enhances injection. The forensic community primarily uses the ABI 3130 XL for the analysis of STRs. Under the analysis parameters typically employed for STR analysis, the amplified DNA fragments must remain denatured [33]. To accomplish this DNA denaturation, the capillary column run temperature is specified to a higher than room temperature, and buffer additives such as formamide, urea, and 2-pyrrolidinone are added to keep the DNA from re-annealing. Even under strong denaturing conditions, DNA molecules can sometimes assume various conformations because of intramolecular attractions and capillary run temperatures are commonly employed to help reduce secondary structure in DNA. Thus, high concentrations of urea and elevated temperatures are used to keep the various STR alleles uniformly denatured, as the mobility of DNA fragments can be affected by its conformation. Even with these measures, a stable ambient temperature, should also be maintained as temperature variations can have profound effects on allele migration.

### The Capillary

The capillary column is pivotal to the separation capabilities of CE. In uncoated capillary columns, residual charges on the silica surface generate a flow of the bulk solution toward the negative electrode. This process known as electro-osmotic flow (EOF) initiates problems for reproducible DNA separations because the velocity of the DNA



molecules can change from one run to the next run. Capillary and microchip channel walls, which contain charged silanol groups, are chemically modified or dynamically coated to prevent EOF in DNA separations [34, 35].

Capillary lifetimes can be improved by rinsing the capillary with simultaneous washes of water, tetrahydrofuran, hydrochloric acid, and polymer solution.

### **Sample detection**

Multiwavelength detection has widened the capabilities of DNA analysis beyond a single-dye color and permitted greater multiplexing for STR markers. The approach to the utilization of this technology is to covalently bind a different dye onto the 5'-(nonreactive) end of each primer or to the set of primers. These dyes have a number of interesting properties. They are excited by a single argon-ion laser tuned to 488 nm, yet fluoresce in different regions of the spectra. A multiwavelength analyzer, like a charged-coupled device (CCD) camera, can then be used to find which dye is present, dependent on the emission of each fragment as it passes the detector window [36]. This technique permits the analysis of fragments of DNA that overlap in size, considering they are labeled with different dyes, which fluoresce at different wavelengths.

## **2.5.5 SAMPLE INTERPRETATION**

### **Software Used**

The ABI 3130 XL data collection software is used to process data from the ABI 3130 XL and produce STR genotypes, performs three main functions: control of electrophoresis run conditions, control of which wavelengths of light will be inspected on the CCD camera through the use of “virtual filters”, and enables sample sheets and injection lists to be created whereby the sample name and processing order are stated [37]. The user inputs the name of each sample and which dye colors are available in a sample sheet. The injection list controls the order in which each sample is injected onto the capillary alongwith the time and voltage for the electrokinetic injection and electrophoresis voltage and run temperature. The virtual filter is also specified in the injection list depending on the dyes present in the sample being analyzed. The output from the data collection program is “raw data” that comes in the form of relative fluorescence units on the  $y$ -axis and amount of data points collected on the  $x$ -axis.

### **The Process**

Sample data collected from the ABI 3130 XL is usually represented in the form of peaks that correspond to the various STR alleles amplified from DNA sample. These peaks are present at various locations in a sample's electropherogram and often plotted as fluorescent signal intensity verses time passing the detector.

The multiplex STR kits in use today take benefit of multiple fluorescent dyes that can be spectrally resolved. The various dye colors are separated and the peaks showing DNA fragments are identified and associated with the appropriate color. The DNA fragments are then sized corresponding to an internal sizing standard. Finally, the polymerase chain reaction (PCR) product sizes for the questioned sample are compared to an allelic ladder that has been sized in a similar fashion with internal standards. The allelic ladder consists of alleles of known repeat content



and is used much like a measuring ruler to correlate the PCR product sizes to the number of repeat units exist for a particular STR locus.

## REFERENCES

1. Forensidnacenter. STR - Forensic DNA Testing System [Serial Online]. 2016 [cited 13 March 2016]. Available from: <http://www.forensidnacenter.com/dna-str.html>.
2. Edwards, A., Civitello, A., Hammond, H. A., and Caskey, C. T. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *American Journal of Human Genetics*, 49(4), 1991, 746–756.
3. Chambers, G. K and MacAvoy, E. S. Microsatellites: consensus and controversy. *Comparative Bio Chemistry and Physiology Part B*, 126(4), 2000, 455-476.
4. Collins, J. R, Stephens, R. M, Gold, B. Long, B., Dean, M. and Burt, S. K. *Genomics*, 82, 2003, 10-19.
5. Weber, J. L and May P. E. *American Journal of Human Genetic*, 44, 1989, 388-396.
6. Urquhart, A, Chiu, C. T, Clayton, T. M, Downes T. Weber, J. L and May P. E. *American Journal of Human Genetics*, 44, 1989, 388-396.
7. Feigelson, H.S., Rodriguez, Robertson, A.S., Jacobs, E.J., Calle, E.E., Reid Y.A. et al. Determinants of DNA yield and quality from buccal cell samples collected with mouthwash. *Cancer Epidemiol Biomarkers*, 10(9), 2001, 1005-1008.
8. Sajantila, A., Puomilahti, S. Johnsson, V and Ehnholm, C. *Biotechniques*, 12, 1992,16-21.
9. Jin, L., Zhong, Y. and Chakraborty, R. The exact numbers of possible microsatellite motifs [letter]. *American Journal of Human Genetics*, 55, 1994, 582-583.
10. Walsh, P. S, Filders, N. J and Reynolds, R. *Nucleic Acid Research*, 24, 1996, 2807-2812.
11. Gill, P., Kimpton, C. P., d'Aloja, E., Andersen, J. F., Bar, W., Brinkmann, B., Holgersson, S., Johnsson, V., Kloosterman, A. D., Lareu, M. V., Nellemann, L., pfitzinger, H., Phillips, C. P., Schmitter, M., Schmeiber, P. M. and Stenersen, M. *Forensic Science International*, 65, 1994, 51-59.
12. Baechtel, F. S., Smerick, J. B, Presley, K. W and Budowle, B. *Journal of Forensic Science*, 38, 1993, 1176-1182.
13. Brody, J.R., Calhoun, E.S., Gallmeier, E., Creavalle T.D., Kern S.E. Ultra-fast high-resolution agarose electrophoresis of DNA and RNA using low-molarity conductive media. *Biotechniques*, 37, 2004, 598-602.
14. Sambrook and Russell. *Molecular Cloning: A Laboratory Manual*, 3rd edition, Cold Spring Harbor Laboratory Press, 2001. (Sambrook and Russell cites the original paper: Warburg, O. and Christian W. "Isolierung und Kristallisation des Gärungsferments Enolase". *Biochem. Z.* 310, 1942, 384-421)
15. Berg, J.M., Tymoczko, J.L., Stryer L. (*Molecular Cell Biology*, 5th ed., WH Freeman, 2002. ISBN 0-7167-4955-6)
16. Milan Bier (ed.) (*Electrophoresis. Theory, Methods and Applications*, 3rd printing, Academic Press, 225. LCC 59-7676, 1959)



17. Santahucia. *Proceeding of the National Academy of Science U.S.A.*, 95, 1998, 1460-1465.
18. Lodish, H, Berk, A, Matsudaira, P, *et al.* (*Molecular Cell Biology*, 5th ed., WH Freeman: New York, NY, 2004)
19. Saiki, R. K, Gelfand, D. H, Stoffel, S, Scharf, S. J, Higuchi, R, Rorn, G. T, Mullis, K. B and Erlich, H. A. *Science*, 239, 1988, 487- 491.
20. Reynold, R, Sensabaugh, G and Blake, E. *Analytical Chemistry*, 63, 1991, 1-15.
21. Birch, D.E, Kolmabin, L. Wong, J. Zangenberg, G. A, Zoccoli, M. A., Mckinney, N., Young, K. K. Y and Laird, W. J, 381, 1996, 445-446.
22. Grossman, P. D., Soane, D. S., *Biopolymers*, 11, 1991, 1221–1228.
23. Innis, M. A, Gelfand, D. H. and Sninsky, J. J (PCR) Application Protocols for functional Genomics. 1999.
24. Moretti, P. R., Baumstark, A. L, Defenbaugh, B. A., Keys, K. M, Smerick, J. B and Budowle, B. *Journal of Forensic Science*, 46, 2001, 647- 660.
25. Lindahl. *Nature*, 362, 1993, 709-715.
26. Pionar, H. N. Progress in Forensic Genetics 9, 2003, 575 – 579.
27. Butlar, J.M., Buel, E., Crivellente, F. and McCord, B.R. Electrophoresis, 25, 2004, 1397-1412.
28. Erfle, H., Ventaki, R., Voss, H., Rechmann, S., Benes, V., Stgemann, J., Ansorge, W. *Nucleic Acids Res.*, 25, 1997, 2229-2230.
39. Berschik, P., Henke, J. Proceedings of the Fourth International Symposium on Human Identification, 1993, pp. 201-204. Madison, Wisconsin: Promega Corporation.
30. Heller, C., Electrophoresis, 22, 2001, 629–643.
31. Grossman, P. D., Soane, D. S., *J. Chromatogr.*, 559, 1991, 257–266.
32. Madabhushi, R. S., Electrophoresis, 19, 1998, 224–230.
33. Klein, S. B., Wallin, J. M., Buoncristiani, M. R., Forensic Science. *Commun*, 5, 2003.
34. Rosenblum, B. B., Oaks, F., Menchen, S., Johnson, B., *Nucleic Acids Res.*, 25, 1997, 3925–3929.
35. Fung, E. N., Yeung, E. S., *Anal. Chem.*, 67, 1995, 1913–1919.
36. Giusti, W. G., Adriano, T., PCR Methods Appl., 2, 1993, 223–227.
37. Genescan Reference Guide: Chemistry Reference for the ABI Prism 3100 Genetic Analyzer, Foster City, CA 2003.