KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY

KUMASI

GHANA

COLLEGE OF SCIENCE

Allele Frequency Distribution of FGA and D18S51 STR Markers among the Asante Population of Ghana

By

Harriet Biaa Agyekum (BSc. Biological Sciences)

A thesis submitted to the Department of Biochemistry and Biotechnology, Kwame Nkrumah University of Science and Technology (KNUST), in partial fulfillment of the requirements for the degree of

MASTERS OF PHILOSOPHY (FORENSIC SCIENCE)

August, 2017

DECLARATION

I hereby declare that this piece is my own work towards the MPhil Forensic Science and that, to the best of my knowledge, it contains no material previously published by another person, nor material which has been accepted for the award of any other degree of the university, except where due acknowledgment has been made in the text.

Harriet Biaa Agyekum	……	…….
(PG 2477714)	Signature	Date
(Student Name & ID)		

Certified by:

Dr. S. Pandam Salifu	……	…….
(Supervisor)	Signature	Date

Certified by:

Dr. Peter Twumasi	……	…….
(Head of Department)	Signature	Date

ACKNOWLEDGEMENT

I am truly grateful to Almighty God for His guidance and how far He has brought me with this thesis. I wish to acknowledge with gratitude the immense assistance offered to me by Professor Kofi Marfo (Professor and Foundation Director, Institute for Human Development, Aga Khan University, Kenya) and his wife Mrs. Rose Marfo. To my parents Mr. and Mrs. Agyekum, my siblings Miss Gloria Takyia Agyekum and Mrs. Abigail Akuamoah-Boateng for their financial support, prayers and encouragement throughout my studies.

I am also thankful to my supervisor Dr. S. Pandam Salifu (Department of Biochemistry and Biotechnology, KNUST), Professor B.W. Lawson (Department of Theoretical and Applied Biology, KNUST) and Dr. Peter Twumasi (Department of Biochemistry and Biotechnology, KNUST) for their guidance throughout this thesis writing. I am grateful to them for his tolerance and substantive advice which has brought this project into being.

Big thanks also goes Dr. Richard Larbi (KCCR, KNUST) and Mr. Bernard Totime (Department of Biochemistry and Biotechnology laboratory) and my friends, Mr. George Aboagye Mensah, Ernest Bonnah, Kwame Obeng and Jeffery Dadson for their assistance and advice during my thesis writing.

Finally, my sincere gratitude also go to Full Gospel Business Men Fellowship (TEK Chapter) members, all my friends and family whose names I cannot mention for their prayers, support and encouraging me in diverse ways during my studies.

ABSTRACT

Short tandem repeats (STRs) are highly polymorphic loci in the human genome and are used worldwide for forensic identification. STR loci or markers are motifs with 2-6 base pair repeats in the chromosome of an individual. A key problem with STR application is that; in most sub-Saharan African countries of which Ghana forms a part, there is limited knowledge and information concerning allele frequency distributions among different populations. In this study, the allele frequencies of 60 unrelated Asante individuals among the Ghanaian population were analyzed using polymerase chain reaction (PCR) and polyacrylamide gel electrophoresis (PAGE). Statistical parameters of forensic importance such as observed and expected heterozygosity (H_O and H_e), homozygosity and polymorphism information content (PIC) were calculated for the study markers FGA and D18S51. The parameters indicated the effectiveness of the markers in forensic study among the Asante population. The PIC, H_e and H_o were 0.8580, 0.8704 and 0.3038 for FGA and 0.8358, 0.8530 and 0.2154 for D18S51, respectively. The obtained information on the FGA and D18S51 markers demonstrates that these loci are useful for forensic and identification purposes. This is the first time to establish the frequency distribution of FGA and D18S51 markers of which it is of much importance to the that Asante population and Ghana as a whole.

TABLE OF CONTENTS

LIST OF PLATES

LISTS OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

DNA – Deoxyribonuclease Acid

STR – Short Tandem Repeat

PCR – Polymerase Chain Reaction

PAGE – Polyacrylamide Gel Electrophoresis

PIC – Polymorphism Information Content

H_o – Observed Heterozygosity

H_e – Expected Heterozygosity

VNTR – Variable Number Tandem Repeats

SNP- Single Nucleotide Polymorphism

SSCP – Single-Strand Conformational Polymorphism

RFLP – Restriction Fragment Length Polymorphism

MtDNA – Mitochondria DNA

MLP – Multi-Locus Probe

SLP – Single Locus Probe

FBI- Federal Bureau of Investigations

FSS – Forensic Science Service

CODIS – Combined DNA Index System

NDIS – National DNA Index System

DEDICATION

I dedicate this work to the Marfo and Agyekum families of Meduma.

CHAPTER ONE

INTRODUCTION

1.1 Background study

One of the most significant current discussions in the legal and moral philosophy that has revolutionized the area of forensics over the past two decades is the discovery of Deoxyribonucleic acid (DNA) typing for human identification (Lynch *et.al.*,2010). The DNA is a highly stable polymer made up of subunits called nucleotides. Humans are made up of 23 pairs of chromosomes which consist of 22 autosomal pairs and a pair of sex chromosomes of which an individual inherits a chromosome from each parent to contribute to the individual's pair (Talwar, 2014; Therman & Susman, 2012).

Many regions of the human chromosomes exhibit diversity among individuals with variable number of sequences and hence they are termed polymorphic meaning "many forms" with the proportion of the DNA that varies from person to person being 0.1%. These polymorphic sequences are used in solving forensic cases, the identification of individuals, paternity testing as well as the diagnosis of genetic diseases (Kashyap *et.al.*,2004).

Central to the entire discipline of DNA is the concept of DNA typing or profiling. Recent developments in the field of forensics such as the popularity of television shows like CSI: Crime Scene Investigation and other movies bordering Law & Order have sparked and led to a renewed interest of the general public in DNA typing. According to Butler (2010), the application of DNA technology to the biological evidence in criminal casework has revolutionized forensic science since the introduction of DNA profiling

and the development of PCR for DNA analysis. This DNA profiling was first introduced by Alec Jefferys, a lecturer in the University of Leicester, UK in 1985.

Most human populations differ based on their genetic composition and these infer credible historical patterns and interbreeding among the populations (Dale*et.al.*, 2006). The ability to identify, with a high degree of certainty a suspect in violent crimes now routinely provides valuable leads to criminal investigators worldwide, often in many circumstances where there are no eyewitnesses. Forensic DNA technology is a very sensitive and universally accepted scientific technique.

The mapping of genetic markers on the human chromosomes and the development of new techniques of polymorphism detection were some of the basic goals of the Human Genome Project (National Research Council, 1996)and a major application of these discoveries was the use of DNA technologies and advancements in the aspects of forensic, which involves the comparisons of DNA profiles of evidence samples with those of one or more known subjects. This helps to identify, narrow down, or exclude the source of origin of the elements of evidence samples (Kirby, 1992; National Research Council, 1996).

Allele frequency is the proportion of a particular allele or variant of a gene among all allele copies being considered. It can be expressed as the percentage of all alleles at a given locus on a chromosome in a population gene pool represented by a particular

allele (King *et.al.*, 2013). The deduction and calculation of allele frequencies is therefore at the heart of our understanding and interpreting of DNA profiles.

According to Lewis (2010), the results of these calculations estimate the probability of an unknown unrelated individual selected at random from a given population having a matching DNA profile. Hence, the allele frequencies serve as reference for determining the significance of DNA evidence in identifying humans.

The Allele Frequency Database (ALFRED) on the other hand has also served as a resource on gene or allele frequency data on human populations and it is supported by the U.S National Foundation. It is designed to make allele frequency data on human population samples readily available (http://alfred.med.yale.edu/). More than 32,200,000 allele frequency tables involving over 663,000 polymorphisms and over 700 populations has been published in this database (http://alfred.med.yale.edu).

Hereafter, the Random Match Probability (RMP) of various alleles which is the probability that an individual chosen at random from a suitable population will have the same DNA profile as that of the evidence sample can then be calculated (Krimsky & Simoncelli, 2013; Lynch *et al.*, 2010). According to National Research Council (1996), the smaller the probability of the random match, the greater the chance that the two DNA samples are from the same person and vice versa (as cited in Pyrek, 2010).

This can be illustrated briefly when both the DNA samples from a crime scene and a suspect are compared, and the two profiles happens to match at every tested locus, then

it may be presumed that the suspect might have left the DNA at the crime scene. Hence, the need to determine the match probability of these loci or the probability of finding the same profile.

In the generation of a DNA profile, DNA fingerprinting is employed. While a variety of definitions of the term DNA fingerprint have been suggested, this paper will use the definition by Roewer, (2013) who saw it as the genetic analysis of biological specimen (blood, semen, swabs, etc.) to generate a DNA profile that can be considered unique to an individual, and hence can be used for identification purposes.

Unlike the traditional fingerprints which can be obtained from the dermal patterns present on the fingertips, DNA fingerprints can be obtained from any part of the human body from the hair down to the toe nails. Although these DNA fingerprints used in the generation of the profiles are unique to every individual just as that of the traditional fingerprints from the fingertips, they are never altered in any way except in situations where mutations might have occurred in an individual at birth (Roewer, 2013).

DNA profiling has been used in generating Forensic DNA databases and are therefore now well established in many countries in the world but there is still less knowledge and information concerning these databases in Ghana. The databases contain DNA profiles of convicted people and the 'risky population' (that is, people who have been arrested before but were innocent and not convicted or people who have been convicted before or received police warnings). Elsewhere, the DNA databases of the whole population may

also be proposed and this helps in tracking relations, reveal more details about a person's health or can assist in medical laboratory tests or that of the courts (Butler, 2011).

Several techniques including Restriction Fragment Length Polymorphisms (RFLPs), Single Nucleotide Polymorphisms (SNPs) and Short Tandem Repeats (STRs) have been exploited by geneticist and forensic investigators in human identification over the years. However, STRs have been proven to be the most acceptable among the rest and as such, widely used in forensic investigations (Kirby, 1992). STR sequence data is expected to increase the effective number of alleles in forensic applications which improves the power of discrimination in many cases (Sun *et.al.*, 2012).

1.2 Problem statement

DNA technology has proved to be a valuable investigative tool, which has helped to exonerate the innocent and to bring those responsible for serious crimes to justice(Hart, 2002; Murangira & Jyoti, 2012). However, a major problem with this kind of application is that; in most sub-Saharan African countries, there is limited knowledge and information concerning allele frequency distributions among different populations due to the non-existence of allele frequency databases for the autosomal STR markers. With such cases, problems can arise with DNA evidence interpretation since DNA collected from a crime scene, which may match that of a suspect will be with no statistical value once the population has no database.

Most studies in allele frequency distributions have been carried out and are therefore now well established in many population in the world such as African Americans, U.S. Caucasians, Hispanics, Far East Asians and Native Americans (Budowle *et.al.,* 2001) as well as some few African countries as among Libyan population (Khodjet-el-Khi *et.al.,* 2012) and some Morrocan populations (Bentayebi *et.al.,* 2014).

Although there are some published works concerning the diversity of Ghanaians, these papers had focused on the X-chromosomal loci and diversity in mitochondrial DNA (MtDNA) (Fendt *et.al.,*2012; Poetsch *et.al.,* 2009; Thiele *et.al.,* 2008).The paucity on the knowledge of allelic frequency and/or genetic diversity within the Ghanaian populations on the whole remains a limitation to the applications of the forensic DNA technologies.

Ghana has no generated allele frequency database to which probabilities of paternity testings' and other forensic cases are compared with for its calculations. DNA evidence is therefore subject to attack in court. An example is that seen in the rape case of R. vs. Musa-Gbengba when the prosecution presented DNA evidence with its RMP based on Afro-Caribbean database instead of using a Sierra-Leonian database. The judge for this case permitted the jury to alter the match probability if they accepted that the figure might have been different if a Sierra-Leonian database was used (McCartney, 2013).

1.3 OBJECTIVES

1.3.1 Main objective

The general objective of this thesis was to assess the allelic frequency distribution of the FGA and D18S51 STR markers among the Asante population of Ghana.

1.3.2 Specific Objectives

Specifically, the study sought to:

- Identify the various polymorphs present at the FGA and D18S51core STR loci within the Asante population.
- Determine the frequency ofeach different allele at the STR loci.
- Determine the frequency distributions of the Asante population based on the FGA and D18S51 markers.

1.4 Justification

There is the need to generate Ghanaian population data sets that provide a solid foundation for allele profile frequency estimates. This data can serve as a reference source for allele frequencies of the studied markers. The STR loci, FGA and D18S51 are the most polymorphic STR marker among most human populations (Butler, 2005). It is expected that these selected STR loci will show much variation within the study population and can therefore be used as an identification tool to help narrow down to individual level of identification.

The Asantes are the largest group among the Akan ethnic group and they are selected for this study because they also happen to be among the largest population in Ghana (Meyerowitz, 1952), hence knowing the allelic frequency of the Asante population will be a step in generating a population DNA data set for Ghana as whole. This can enhance solving of forensic and identification issues on the Asante population, not only to solve crime but also to serve as a basis for the establishment of genetic relations especially paternity.

CHAPTER TWO

LITERATURE REVIEW

2.1 DNA and its structure

Deoxyribonucleic acid (DNA) can be defined as a molecule that carries all the hereditary genetic information of an organism for its development, living and reproduction (Amatucci, 2010; Jutta, 2013). It is the blueprints for an organism's make up (Brown & Botstein, 1999). For humans, it carries all the genetic information that make us humans (Amatucci, 2010; Donoghue, 2013). It is made up of repeating units of nucleotides containing a sugar, a phosphate group and a nitrogenous base with the phosphate of one nucleotide connecting to the sugar of another (Egli & Saenger, 2013; Sinden, 2012).

According to Moseley (2011), a DNA strand consist of two main bases which are the purines and pyrimidines. However, these bases are distinct in their constituents with purines making up of guanine (G) and adenine (A), while the pyrimidines are cytosine (C) and thymine. There is a complementary and rudimentary base pairing between the nucleotides of which (A) pairs with (T) and (G) pairs with (C) (Lynch *et al.*, 2010; Rettner, 2013). These base pairs attach themselves to a sugar (pentose sugar) and a phosphate group leading to the formation of a chain of nucleotides. The DNA's instructions or genetic code is controlled by the order and arrangements of these bases (Rettner, 2013).Human DNA is made up of about 3 billion bases with over 99 percent of them being the same in all humans (Jutta, 2013). The DNA is mostly found in the nucleus of the cell but may also be located in the mitochondria (National Institutes of

Health, 2015). It is usually a double-helix with the two strands running in the opposite direction. One strand of the DNA leaning in the 5' to 3' direction and the other in the 3' to 5' direction (Sinden, 2012; Tamarin, 2015).

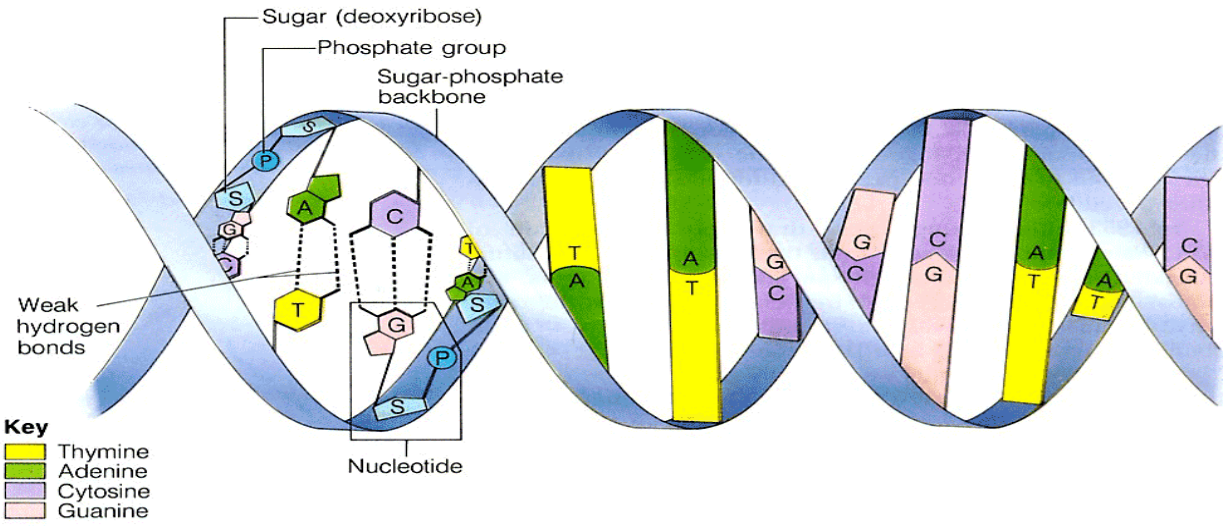


Plate 2. 1 A diagram of the DNA structure showing the base pairs, sugar and phosphate backbone (http://quotesgram.com/dna-structure-quotes/).

2.2 Organization of DNA in the cells

The human DNA plays two important roles which includes the carrying of instructions used in making the components of a cell and providing ways for these set of instructions to be passed on to the daughter cells as cell divisions occurs. In eukaryotes, DNA is packaged into chromatin whose fundamental repeating unit is in the structure of the chromosome (Yazdi *et.al.*,2015;Jiang *et al.*,(2013) pointed out that, in at least every two unrelated individuals, 99.9% of their genomic DNA sequence are shared and the remaining 0.1% accounts for the minor differences in the DNA. These lead to a lot of

physical differences such as race, eye, hair and skin colors in human (Alexov, 2014; Jiang *et.al.*, 2013). A considerable number of literature has been published stating that all individuals are genetically different with the exception of identical twins (Donoghue, 2013; Edwards, 2013; Noyd *et.al.,* 2013; Rafferty, 2011) and that these differences are due to the various polymorphisms or variations in the DNA sequence of each person.

There are also proteins related to the chromosomes with ordered sequences that bear portions of the hereditary information concerning an organism and these are called genes (Panayiotopoulos 2010) which determine these physical differences or genetic traits. These portions of the gene that codes for proteins are called the exons whereas the introns are portions of the gene that do not directly code for proteins.

The locus (plural; loci) which is the specific position of a gene or DNA sequence on the chromosome can vary based on the different forms of the gene or the DNA sequence (Butler, 2005; Butler, 2009; Panayiotopoulos, 2010). These variant forms of the gene or sequence is what is referred to as "alleles" (Crowley, 2011; Daniels & Nicoll, 2011). Crowley (2011); Daniels &Nicoll (2011) mentions that humans always have two alleles at each genetic locus which can be either homozygous (having the same forms of alleles) or heterozygous (having different forms of alleles).

2.3 Genetic polymorphisms

The portions of the genome where there are enough variations to determine the diversity among individuals are called the polymorphic regions (David, 2005). These regions are mostly within the non-coding regions and are used for forensic analysis due to that fact that they make up approximately 95% of the genomic DNA and do not code for proteins (David, 2005). These variations in the non-coding region of the DNA sequences accounts for the uniqueness of individuals alongside their genetic relatedness (Lach, 2006).

Variations among individuals are basically dependent on mutations that occur in the DNA sequence. These mutations arise as a result of altered DNA sequences which may be due to either change in the base sequence or change in the number of copies (alleles) of a particular DNA sequence (Strachan *et.al.,* 2014). Variations that comes in the form of single base changes are the single nucleotide polymorphisms (SNPs) (Alkan *et.al.*, 2011; Ferrier, 2013).

SNPs are single base pair positions within the human genomic DNA where variants of alleles or alternative forms of alleles exist in individuals within a population and can be detected by sequencing, PCR, RFLP or Single-Strand Conformational Polymorphism (SSCP) methods (Kashyap *et.al.*, 2004). For example SNP may result in the sequence GATA changing to GGTA. In about every 1000 base pairs of the human genome, SNPs may be found (Robert *et.al.*, 2014).

There are also some particular core sequences in the DNA that repeat several times in the DNA sequence leading to polymorphisms (Jobling *et.al.*, 2013). These sequences are referred to as tandem repeats, which are used in forensic DNA analysis. These tandem repeats may range from 2 to 100bp and are categorized on this basis as variable number of tandem repeats (VNTRs) and short tandem repeats (STRs).

VNTRs are common hypervariable regions used in genetic identification and testing of relations due to their high polymorphisms (David, 2005). They are regions of the human DNA composed of several hundreds of base pairs arranged as tandem repeat units, which vary among individuals. They are mostly made up of about 8-80 base pairs and are also known as minisatellites (Jeffreys *et.al.*, 1985).

Short tandem repeats (STRs) on the other hand are repetitive elements comprised of repetitive DNA motifs of 2–6 bp in length (Ellegren, 2004) and with various repeats which vary between or among individuals. The number of repeats can vary greatly between individuals at a specific locus, henceforth the analysis of several STR loci together produces information that can be used to profile an individual. STRs are also multi-allelic and produce many possible genotype combinations that can aid DNA mixture interpretation (Butler *et.al.*, 2007). Again, its higher variability has made them excellent markers of choice in forensic DNA analysis for the past twenty years in large measure for human identification (Butler, 2005).

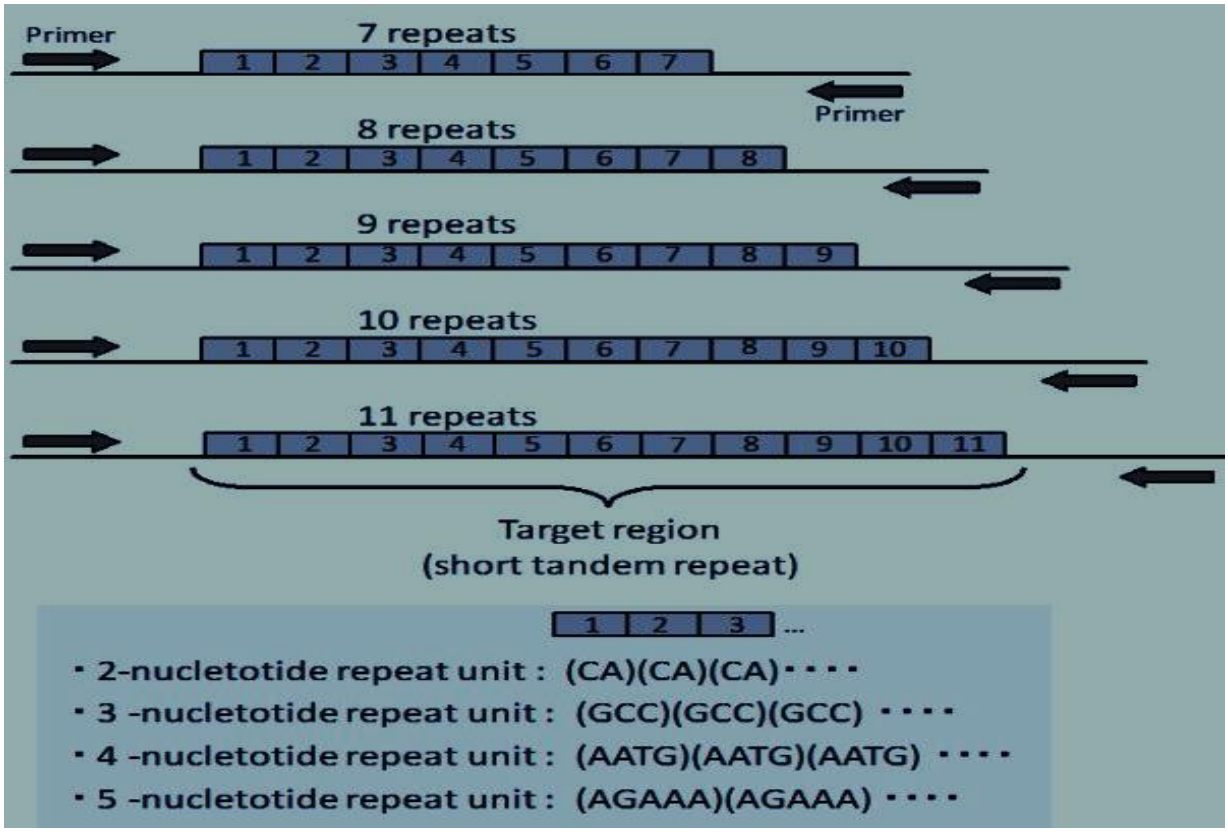


Plate 2. 2STRs showing various number of alleles.

(http://www.intechopen.com/books/biometrics/dna-biometrics).

STR loci may be classified based on the number of allele repeats, the unit length or the number of nucleotide repeats. They may be dinucleotides, trinucleotides, tetra nucleotides and penta nucleotides having two, three, four and five nucleotide repeats respectively. Tetra nucleotide repeats are the most common and popular STR markers used in forensic fingerprinting or investigations (Butler, 2005).

STR loci may be grouped into simple, compound or complex repeats based on how the alleles conformed to the core repeat pattern. Simple repeats have their repeats composed of the same sequence and length, and include TPOX, CSFIPO, D5S818, D13S317 and D16S539. There are others STRs with simple repeats but with non-consensus alleles and these include TH01, D18S51 and D7S820. Compound repeats are repeats made up of two or more simple repeats which comprises of vWA, FGA, D3S1358 and D8S1179. D21S11 is categorized as an STR with a complex repeat sequence having multiple repeat units of various lengths along with several intervening sequence (Butler, 2005).

Among all these STRs, FGA, D18S51 and D21S11 are the most polymorphic loci within various populations with TPOX being the least polymorphic. Over the years, the human identity testing has settled on the use of core STR loci that are widely used in DNA profiling applications (Butler, 2006) and have been described as very effective tools for human identity testing in the early 1990s (Edwards*et.al.*,1992). STRs has grown popular because of its advantages over the use of other DNA analytical techniques such as the use of VNTRs or minisatellites which are older generation of DNA markers used in forensic DNA analysis.

DNA evidence recovered from crime scenes is often degraded with long DNA molecules or strands being broken into smaller fragments. In such instances, STRs are much intact and suitable to use for DNA analysis compared to VNTRs and this is because STR alleles are generally shorter than the VNTRs having about 100-400 bp whereas the VNTRs have about 400-1000 bp. Additionally, the usage of STR markers

tends to improve PCR amplifications since smaller alleles have the tendency to amplify better than larger alleles (Butler, 2005).

According to White*et.al.*,1997, most STR markers are analyzed as DNA fragments made up of about 100-350 base pairs, having repeats units of 2,3,4 or 5 bases with their electrophoretic discriminations based on the sizes and distribution of the repeating units for the different alleles present together with the specificity and efficiency of polymerase chain reaction (PCR) as a determinant of quality amplicons. STRs occur approximately once in every 10,000 bp and is represented by about 3% of the human DNA (Osman *et al.*, 2015; Shrivastava *et.al.*, 2015). DNA samples are characterized by the size of the STR fragment detected and the identification of alleles at each locus is the number of times a repeated unit is present within an identified section or STR fragment. Most STR loci selected for forensic use generally has about 7 to 30 different alleles on the average (Kashyap *et.al.*, 2004).

2.4 FORENSIC DNA ANALYSIS

2.4.1 DNA fingerprinting

The first method of DNA testing and analysis was DNA fingerprinting, which was carried out using restriction fragments (Linacre & Tobe, 2013). DNA fingerprinting is the comparison of evidence samples established from crimes scenes with that of a suspect's using various techniques such as restriction enzyme digestions, southern

blotting and autoradiography (Lynch *et.al.*, 2010) for identification of individuals and paternity testing (Sergio, 2013).

Several developments in molecular biology including restriction enzymes, Sanger Sequencing and Southern blotting were advanced in 1960s and 1970s which helped scientist in examining DNA sequences (Goodwin *et.al.*, 2011). In 1978, DNA polymorphisms were also identified using Southern blotting after which the first highly polymorphic locus was identified and reported in 1980 until in September 1984 when the forensic application of DNA was first discovered and termed as DNA fingerprinting by Sir Alec Jefferys at the University of Leicester in UK (Goodwin *et.al.*, 2011).

Restriction Fragment Length Polymorphisms (RFLPs) were initially used in analyzing single-base pair polymorphs and minisatellite VNTRs, in which the loci varied in length based on the restriction fragments. It is the earliest method of forensic DNA analysis that involved the comparison of lengths of specific DNA fragments of evidence samples in which DNA is subjected to various restriction endonucleases (enzymes) such as Hind III, Bam H1, Eco R1, etc. to cut the DNA into various fragments of interest. This technique is used in the analysis of various polymorphisms or differences between multiple samples (Butler, 2001).

RFLP was used in the early 1980s when DNA fingerprints was first used in establishing the paternity of some children and it became the most dependable tool for paternity testing in the history of genetic variation studies since it had an exclusion power of

99.99% (Butler, 2008). However, this technique requires high amount of non-degraded DNA, it is labor-intensive and time consuming as well (Primorac *et.al.,* 2000).

Before the use of RFLP techniques where restriction enzymes were used in cleavage, multilocus probes (MLPs) which also consisted of "core" sequences were simultaneously used for identification of polymorphs of sequences at various loci (Butler, 2005). However, the profiles generated by these probes were complex and could not be used to identify alleles at different loci and this brought about the use of single locus probes (SLPs) which had unique DNA sequences flanking the VNTRs under study. Henceforth, these SLPs were used in the characterization of various polymorphs at a single locus for the detection of individual alleles (Kashyap *et al.*, 2004).

According to Jeffreys *et al.*, (1985), he identified hypervariable loci called minisatellites which were detected through the hybridization of probes to southern blots of restriction enzyme digestions of the genomic DNA. In several publications, Jefferys and his colleagues also reported the techniques of DNA fingerprinting by the use of DNA hybridization probes which composed of VNTRs of core nucleotide sequences. By the use of southern blotting, they detected multiple variable human DNA fragments. VNTRs were then introduced into forensic DNA fingerprinting between the late 1990s and early 2000s as a technique for human identity testing (Budowle *et.al.*, 1991).

Between 1985 and 1986, DNA fingerprinting was applied to real cases where it was used to resolve an immigration dispute by the United Kingdom Home Office as well as

in convicting Colin Pitchfork in 1988 as the killer of two school children in Leicestershire, UK. It was in this same year, 1986 that Cellmark and Lifecodes started the use of multilocus probes for DNA analysis. The Federal Bureau of Investigations (FBI) in 1988 for some technical reasons implemented the use of single-locus probes to score multiple VNTR loci by repeated re-hybridization using southern blotting (Butler, 2005) and in the same year, the UK Home Office and the Foreign and Commonwealth Office questioned the use of DNA fingerprinting in solving family relation issues and the authentication of relationships in immigration cases.

The application of DNA fingerprinting therefore came under serious inquiry in court proceedings due to its usage in parentage testing and forensic investigations which brought about more debates over the statistical strength of DNA evidence in the identification of individuals (Chakraborty & Kidd, 1991), hence the need for STR DNA analysis.

2.4.2 STR DNA profiling

STRs have been widely used in different aspects of modern scientific research, such as human evolution, forensics, anthropology, and disease development (Blessmann *et al.*, 2008; Calafell *et.al.*, 1998; Divne *et.al.*, 2010; Lewis, 2010) in many countries across the globe. Therefore, it is important to profile the STRs for local populations, which are helpful for further genetic characterization.

In Libya for instance, STR markers have been used to describe the various polymorphisms of some selected STR loci to determine the diversity of the general Libyan population to be used in resolving forensic cases (Khodjet-el-Khil *et.al.*, 2012). Forensic genetic parameters of interest such as the Power of Discrimination (PD) and Probability of Exclusion (PE) values were greater than 0.999 which showed significant variations among the Libyan population when compared to several data from other North African populations such as Tunisians, Egyptians and Moroccans. Khodjet-el-Khil *et.al.*, (2012) therefore concluded that, the 15 selected STR markers for the Libyan population had established their utility for forensic and paternity investigations due to the high discrimination detected.

Since STRs are highly polymorphic and informative genetic markers and hence provides a higher individualization and discriminating power, they are well accepted, documented and applied in the forensic setting. They are therefore encouraged by many research communities for their proper use in human identification (Budowle *et.al.*, 2001) and used in the generation of population genetic database which provides estimates of the frequencies of DNA profiles. From Biesecker *et al.*, (2005) and Mpoloka *et.al.*, (2008), STRs have demonstrated very high degree of variability among studied populations and have therefore shown that they can be used in identification procedures. Yasin *et al.*, 2005 also demonstrated this among African Jordanian population using 15 STR markers.

2.4.3 STR Development

For the effective use of STR loci in DNA profiling over a wide range of jurisdictions, a common and standard set of STR markers must be used (Butler, 2015). The common STR loci currently in use were established and categorized initially by the Forensic Science Service (FSS) in England.

With the advent of time, Promega Corporation in the beginning commercialized most of the STR markers developed by the Caskey laboratory. The Applied Biosystems also commercialized the markers of the FSS together with some new markers. Presently, both Promega and the Applied Biosystems have commercial STR kits that constitute common loci for use by the DNA profiling community. There are also various multiplex amplification STR kits that has been introduced in the analysis and profiling of forensic DNA sample (Butler, 2011).

The "first-generation multiplex" consisting of TH01, FES/FPS, vWA and F13A1 had a matching probability of about 1 in 10,000. The FSS also introduced the "second-generation multiplex" (SGM) which also included six STR loci and the sex-determining loci (amelogenin). The six STR loci were TH01, FGA, vWA, D8S1179, D18S51 and D21S11. These markers were very polymorphic with a match probability of about 1 in 50 million.

These common standardized STR markers has been used for DNA typing to make them very effective in identification of individuals across the globe. In November 1997, the laboratory of the Federal Bureau of Investigation (FBI) also set some STR markers to be used as core 13 STR loci within the United State and were used in the generation of the national DNA database (NDIS, National DNA Index System).

Initially, these 13 core United State STR loci were used for Combined DNA Index System (CODIS) to facilitate identification in the States. These loci included FGA, CSFIPO, TH01, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51 and D21S11. Out of these thirteen (13) STR loci, eight (8) of them overlap with the loci and were compiled in the United Kingdom and other European nations. These are FGA, vWA, D3S1358, D8S1179, D18S51, D21S11, TH01 and D16S539 (Butler & Hill, 2012).

According to Cortellini & Cerri (2011), the European Standard Set (ESS) of STR loci was originally made up of seven loci which comprised of FGA, TH01, vWA, D8S1179, D18S51, D21S11 and D3S1358. The European Network of Forensic Science Institute (ENFSI) in April 2009 selected five additional STR loci to be added to the already European Standard Set (ESS). These five (5) included D12S391, D1S1656, D2S441, D10S1248 and D22S1045 (Butler, 2001; Cortellini & Cerri, 2011).

The FBI laboratory also expanded the U.S set core loci to limit the likelihood of adventitious matches with the increasing number of profiles stored within the national

database to expand the international compatibility to assist in law enforcement data-sharing efforts and also to increase the discrimination power to help resolve missing person cases (Hares, 2012).

The expansion now consist of twenty-one (21) set loci and these are FGA, CSFIPO, TH01, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, D12S391, D1S1656, D2S441, D10S1248, D2S1338, D19S433, Penta E, DYS391 and Amelogenin (sex-typing marker). These markers are therefore used extensively in many population studies for forensic purposes and mostly in the generation of allelic profiles.

2.4.4 Application of DNA profiling

DNA databases have become a very significant tool in forensic investigations of individuals mostly criminals since the identification of these individuals require reference samples for comparison. These reference samples are already stored databases generated from biological materials or evidence left on personal items such as brush, combs, razors (Butler, 2001), dresses, body fluids, organs and tissues of perpetrators and other crime scene evidence samples. They may also be obtained from surviving relatives in the cases of missing persons or mass and natural disasters. For other instances, case probabilities based on already compiled DNA databases of the specific population may also be used as reference to which the DNA of victims and suspects are compared with

(Biesecker *et.al.,* 2005). As more samples are added to these databases, there is a continuous increases in the probability of case-case match or case-convicted offender match (Butler *et.al.,* 2004).

DNA profiling is therefore used most extensively in many forensic investigation including both criminal and civil cases to help solve such cases. It is mostly used by the police to help identify suspected criminals or perpetrators of crime and in the same way, helps in exonerating the innocent. It is also used in determining paternity disputes, identification of individuals in missing persons or victims of natural disasters or mass destruction (Diallo, 2013; James, 2012; Jobling & Gill, 2004; Lynch, 2012; Panneerchelvam & Norazmi, 2003).

In a cases where two young girls had been raped and murdered in two nearby English villages, semen collected from the two crime scenes was able to help exonerate a suspect (Richard Buckland) and to finally convict the actual perpetrator (Colin Pitchfork) who happened to have persuaded his friend to give a sample on his behalf when samples were being taken from possible suspects. Pitchfork was arrested and convicted after confessing that he was the actual criminal when he was confronted with evidence that his DNA profile or fingerprint matched the trace evidence collected from the crime scenes (Roewer, 2013).

It is also being used in solving many identification issues as seen in mass and natural disasters. For instance, DNA fingerprinting was used in identifying victims who were

involved in the September 11, 2001 World Trade Centre attacks in the USA where about 850 out of the 1594 victims were solely identified based on DNA fingerprints (Biesecker *et.al.,* 2005). It was DNA analysis that was applied in the Persian Gulf War when STRs were first used in the identification of human remain in 1991 (Kashyap *et.al.,* 2004).

For paternity testing, DNA fingerprinting is extensively utilized and was first used officially in 1985 as an investigating tool in an immigration case. It was used to demonstrate that a boy who was suspected of trying to join his mother in Britain with a fake passport was really the actual child of the woman in question. The analysis also showed that all the biological children of this woman had the same biological father (Goodwin *et.al.*, 2011).

STRs have dominated the genetic information gathered on humans till date due to the use of these STR loci in the U.S national database. Other nations and other criminal justices across the globe have begun to develop National databases. Over five million profiles are now in existence in DNA database that contain information about the loci for forensic applications (Butler, 2005). In the U.S., DNA databases begun when the DNA identification Act of 1994 was passed which empowered the Federal Bureau of Investigation (FBI) to conduct a pilot study into a national DNA database using the Combined DNA Index System (CODIS) as a tool for solving criminal cases.

The use of the CODIS blends the analysis of DNA with computer technology to empower various U.S. crime laboratories at the local, state and national levels to

exchange and make comparison of DNA profiles electronically (McEwen, 1995). These databases based on the thirteen (13) core STR loci and the amelogenin (X and Y) marker constitutes the Forensic Index (made up of DNA profiles from databases) (Moretti *et.al.*, 2001).

2.4.5 Stages of DNA profiling

The main stages of DNA profiling includes extraction of DNA, quantifying the extracted DNA, STR amplification of quantified DNA (PCR), electrophoresis (gel or capillary) and interpretation of data.

2.4.5.1Extraction of DNA

DNA is extracted from biological materials. Human DNA may be obtained from blood, sperm, semen, vaginal fluids, mucus, sweat, saliva, hair roots, earwax, bone, teeth, cerebrospinal fluid, many body fluids either in the dried or wet forms and other body organs such as the skin, muscles, (Butler, 2011; Gardner & Anderson, 2015; Singleton, 2012). It may also be found on cigarette butts, door bells, clothing, credit cards (ATM cards), gloves, hammer, mobile phones and bottles (Shewale & Liu, 2013) when these items were in contact with biological materials.

The extraction of DNA from biological samples involves the collection of cells containing the DNA, lysis, protein digestion and precipitation of the DNA. The more the

amount of DNA initially collected, the more the DNA is precipitated (Hearn & Arblaster, 2010). During the DNA lysis, the sample is treated with lysis buffer which breaks open the cell membrane and nucleus to release the DNA. Detergents such as the Sodium Dodecyl Sulfate (SDS) may be used in the lysis. Treatment of the DNA with proteinase K releases the DNA from DNA associated proteins (histones) and other cellular proteins. These proteins are then precipitated by the addition of some salts such as sodium acetate or ammonium after which the DNA is also precipitated by mixing it with cold ethanol or isopropanol and then centrifuged.

There are a number of methods for DNA extraction which may depend on certain factors such as the nature of the sample (Taupin, 2013), the type of sample, the quantity or the ability to automate the extraction procedure. Some of these methods includes the Chelex 100 method, Silica based method (Goodwin*et.al.*, 2007) and the Phenol-Chloroform method. According to Harty *et.al.*, (2000), blood samples provide a higher yield of quality DNA compared to other biological samples.

2.4.5.2 DNA Quantification

After the isolation of DNA from a sample, DNA quantification is required to ensure that the extracted DNA is from human source. The isolated human DNA can be present with other sourced DNA such as bacteria, plants, animals and fungi (Butler, 2011; Taupin, 2013). The main purpose of DNA quantification for forensic analysis is therefore to identify the suitable amount of DNA to be used as template for PCR amplification and to

prevent the generation of artifacts and other off-scale data such as loss of some alleles (allelic dropout or null allele) and stutter products that make interpretations difficult, challenging and time consuming.

Again the current DNA profiling methods such as STR analysis, MtDNA sequencing and the Y-STR Chromosomal analysis demand very accurate amounts of DNA for processing (Gall, 2011; Mozayan, 2010; Taupin, 2013). In the analysis of STRs, DNA quantification is important in PCR amplification since smaller concentration ranges of the DNA works best and even much better with the multiplex STR analysis. For most commercial kits, a range of 0.5ng to 2.0ng of human DNA is ideal for amplification. The amount of DNA must therefore be attuned and normalized when the amount of DNA falls below or beyond the required range (Butler, 2011). This can be achieved by diluting or concentrating the sample to the desired concentration or amount. DNA quantification saves time during data interpretation as whatever result that is obtained will be accurate (Butler, 2011).

Several methods are employed in the quantification of DNA for the estimation of the amount of DNA present in the sample (Butler, 2009; Li, 2008; Mozayan, 2010) which includes spectrophotometry (UV absorption), ethidium bromide fluorescence emission (Barbas *et.al.*, 2007), real-time PCR (Mozayan, 2010; Shewale & Liu, 2013). There are commercial quantification kits also available for the quantification of DNA.

2.4.5.3 STR Amplification

The next stage after extraction and quantification of DNA for DNA profiling is the amplification, which is a PCR-based technique. The process generates more copies of the DNA at each STR loci being analyzed by targeting STRs on particular chromosomes that are variable with specific primer sequences (Taupin, 2013). The PCR technique is well suitable for the DNA analysis of forensic samples compared to the conventional RFLPs because it is more sensitive, rapid and produces quality amplicons to be detected (Butler, 2011).

The amplification of the STRs are mostly by a pair of primers flanking the tandem repeats which produces resulting amplicons with varying lengths depending on the repeat unit present for each allele. For heterozygote alleles with two different repeat motifs, two different sized amplicons are produced with one amplicon being produced for homozygote alleles having the same repeat number for both alleles.

During STR analysis, more than one STR loci can be amplified at the same time. This simultaneous amplification of more than one STR region on different chromosomes by the use of different primer sequences for each loci is termed multiplex PCR or multiplexing (Butler, 2011).

In the 1990s, STR megaplex kits were produced that comprised of up to eight (8) STR markers which could be amplified simultaneously in a single reaction and detected using fluorescent dyes. Off late, most forensic laboratories use multiplex PCR kits containing sixteen (16) loci which are made up of the thirteen (13) CODIS based on the manufacturer's preference. Some other non-CODIS STR markers and the amelogenin or sex determining marker found on the X and Y chromosomes are also used (Ubelaker, 2012).

2.4.5.4 Electrophoresis

Electrophoresis is the separation of DNA fragments according to their respective sizes due to their movement through a medium with an electric field applied. Since DNA is negatively charged, it migrates towards the positively charged (anode) when placed in an electric field (Pláteník, 2009).PCR products can be separated on the basis of difference in length which is as a result of the variations in the number of tandem repeats using electrophoresis. The variant alleles present at a locus moves through the separating medium at different velocities due to the influence of the applied voltage and are detected as bands or peaks at the end of separation depending on the method being used. These bands or peaks may be visualized by staining, autoradiography or transferred to a membrane surface for detection by other methods.

The commonly used electrophoretic methods for STRs are the slab gels (gel-based electrophoresis which includes agarose gel and polyacrylamide electrophoresis) and

capillary electrophoresis (Hosler & Murphy, 2014; Kobilinsky, 2011). DNA separations by size-based using gel electrophoresis was traditionally used for STR analysis and were visualized on agarose gels or silver stained polyacrylamide gels until the recent introduction of technology which is based on fluorescence detection methods like the capillary electrophoresis (Butler, 2012).

Although the use of agarose and polyacrylamide gels electrophoresis is a standard method for separating, identifying and purifying nucleic acids due to the porous nature of the gels, capillary electrophoresis provides the most reliable method to probe or analyze STRs. It provides better specificity, sensitivity and accurate sizing of PCR products for detection (Hosler & Murphy, 2014).

With the use of capillary electrophoresis, fluorescent dyed markers are used for the STR amplification. The PCR product samples are mixed with denaturing solutions that separates the DNA strands. The samples are heated and loaded into the instrument (mostly called genetic analyzer). This is run at a lower voltage which draws a very small portion of the sample into a capillary which is a thin tube filled with polymer in the instrument. Upon the application of the voltage, the DNA molecules migrate through the polymer solution with the smaller ones moving faster and resulting in the separation of alleles. At the end of the capillary tube, the migrants passes through a tiny window where a camera detects the fluorescence of the primers and collects the color and intensity data (peaks) (Butler, 2001).

Polyacrylamide gels on the other hand constitute cross-linked chains formed by polymerizing acrylamide with a cross-link agent called N,N-methylenebisacrylamide (Erber, 2010). This reaction is carried out with ammonium persulfate (APS) as the initiator and TEMED (N,N,N,N-tetramethylethylendiamine) as the reaction catalyst (Barril & Nates, 2012). Polyacrylamide Gel Electrophoresis (PAGE) is used in the separation of shorter nucleic acids or DNA fragments ranging between 1-1000 base pairs depending on the concentration of the acrylamide used. It has a higher resolution power, can accommodate much larger quantities of DNA and produce very pure DNA compared to agarose gels (Barril & Nates, 2012; Guilliatt, 2002).

Staining of polyacrylamide gels with by silver staining is suitable for detection of PCR products though other stains such as SYBR Green or Ethidium bromide may be used, hence used comprehensively in imagining DNA. The process involves the transfer of the gel comprising the separated bands into a tank filled with solutions that tends to uncover the DNA bands to the stain.

2.4.3.5 Interpretation of Data

After electrophoresis, the peaks **(plate 2.3)** or bands (**plate 2.4**) of the separated DNA fragments detected are compared to standard DNA sizes and or allelic ladders of the various loci being analyzed. These peaks or bands aid in the interpretation of the profile obtained. The length of the DNA from each band or peak are predictable by matching

the distance travelled to a standard DNA or ladder that has travelled the same distance (Kobilinsky, 2011).

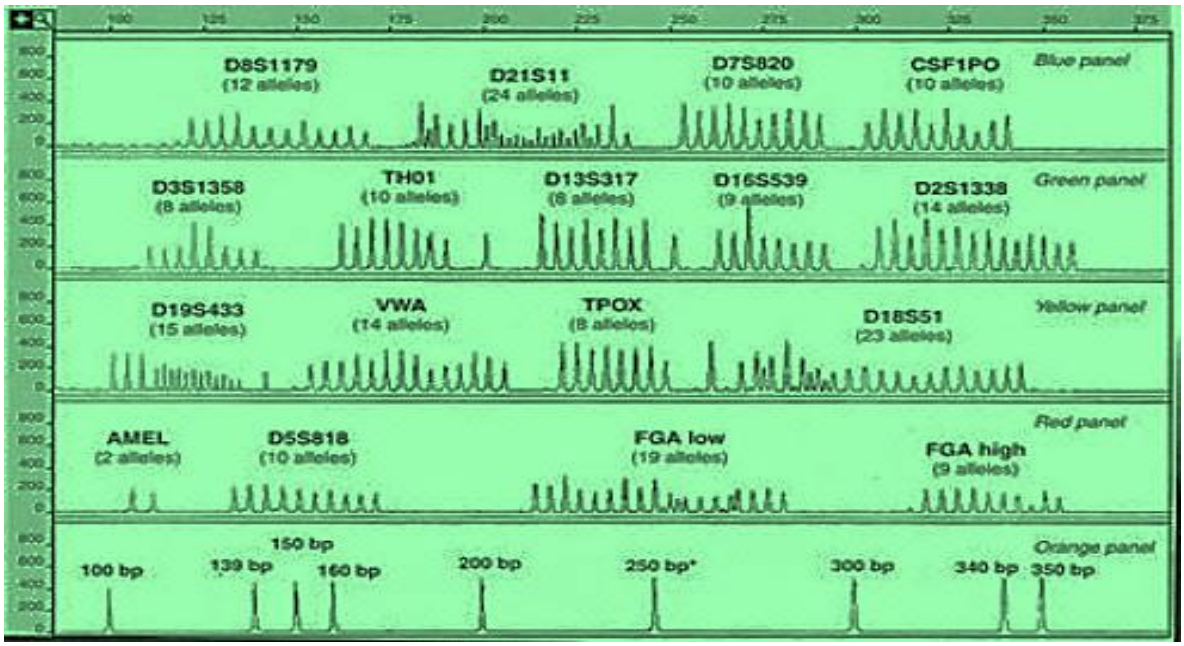


Plate 2.3 Alleles of 15 STR loci and the amelogenin sex-typing test from the AmpFISTR Identifiler kit.

The bottom panel is a "sizing standard"—a set of peaks from DNA sequences of known lengths (in base pairs). The numbers in the vertical axis in each panel are relative fluorescence units (RFUs) that indicate the amount of light emitted after the laser beam strikes the fluorescent tag on an STR fragment(*Reference Manual on Scientific Evidence*, 2011).

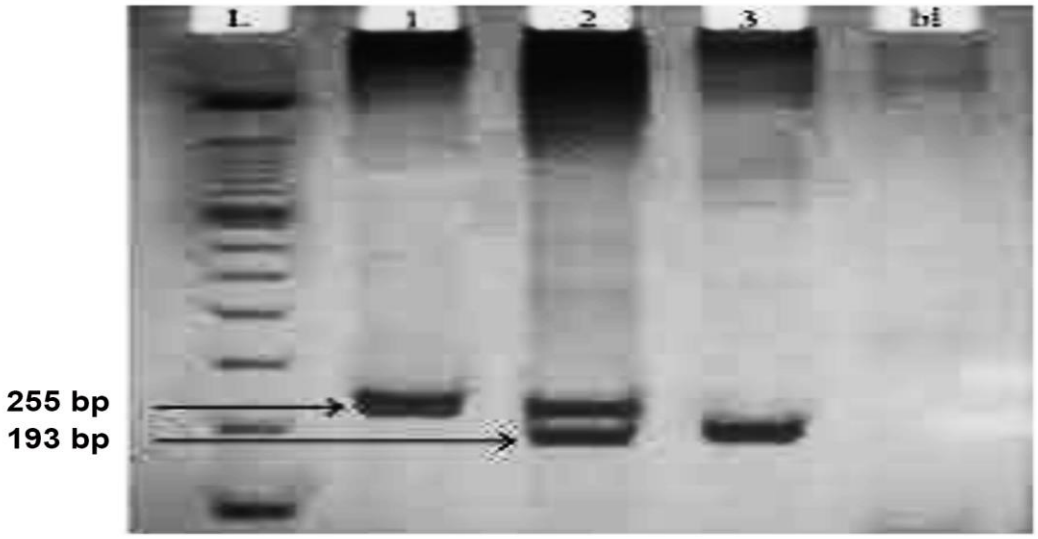


Plate 2.4 The PCR products (bands) detected using Silver Staining Method after Polyacrylamide Gel Electrophoresis.

Lane L: DNA Ladder 100 bp; lane 1: CCR5 wild-type homozygous genotype (CCR5/CCR5, 225 bp); lane 2: heterozygous genotype (CCR5/Δ32, 225 bp, and 193 bp); and lane 3: variant allele homozygous genotype (Δ32/Δ32, 193 bp); bl represents blank reaction (without DNA) (http://www.hindawi.com/journals/ah/2014/924030/fig1/).

2.5 Allele Frequency Database

Various researches and studies (Budowle *et.al.*, 2005; Goodwin *et.al.*,2001; Yasin *et.al.*, 2005) have shown the genetic variations based on allele frequencies among most populations and within many sub-populations arising from the existence of various alleles at different loci. All genetic markers shows some specific frequency occurrence within a population (Rudin & Inman, 2001), hence the need to identify the frequency of a marker's occurrence so as to attach significance to any peculiar trait. Allele frequency database are generated allele frequencies for a particular populations which are compiled from unrelated persons within the said population (Butler, 2001).

The STRBase located at http://www.cstl.nist.gov/biotech/strbase/was initially compiled and launched by John Marshall Butler in the late 1996 and early 1997 at the National Institute of Standards and Technology (NIST). This is an internet database with its sole aim of bringing together the abundant literature and information about STR markers to make current and future forensic DNA works much easier. Most of the fundamental materials, resources and information came from his PhD dissertations and that of other researches from various article in forensic journals and other sources (Butler, 2008).

The version 2.0 of the STRBase generated by the European Network of Forensic Science Institutes DNA Working Group (ENFSI DNA WG) for instance stores data of various alleles studied and can be used in the calculation of the match probabilities of each STR allele within various Caucasian populations across Europe. It provides the frequencies of all the alleles present for a particular locus within a specific Caucasian population (http://strbase.org/about and http://strbase.org/frequencies).

In Botswana, the first attempt to determine the allele frequency of some five sub-populations in the country which could be used in developing a national allele frequency database was made. The study included vWA, FGA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D18S51, D21S11 and Amelogenin (sex determining locus) (Mpoloka *et.al.*, 2008). All these alleles showed very high degrees of polymorphisms with their heterogenisity above 67% and a match probability of less than 0.2, suggesting that there was no significant inbreeding within the sub-populations studied and a higher

discriminating power. Such frequency of alleles from the studied populations can be used for identification of individuals in Botswana.

Studies were also conducted for the first time in Southern Iraq on 100 unrelated individuals from the Mesan and Basra provinces, where genotyping of some selected STR loci was done to generate the allele frequency and for that matter a genetic database for the population (Hameed *et.al*., 2015). The heterozygosity and Power of Discrimination (PD) were found to be 0.696 and 71%-97% respectively, indicating a higher genetic diversity among the population and can be used in generating an allele frequency database for the Iraq population (Hameed *et al.*, 2015).

2.6 Genetic Diversity Studies on Ghanaian Populations

Few works has been published concerning the diversity of Ghanaian populations with most of them focusing on the diversity of the X-chromosomal STR loci within the population (Poetsch *et.al.*, 2009). Eleven (11) typed X-chromosomal STRs situated in four different linkage groups from 129 men and 114 women who were unrelated Ashanti's using two multiplex PCR methods were used to typed. A study on the Ewe population of Ghana also calculated the allele frequency of 8 X-chromosomal markers within the population and compared these frequencies to known data from Germans and Amharic population in Ethiopia (Thiele et.al., 2008).

Another study on 191 Akan people where their full mitochondrial control regions were sequenced (Fendt *et.al*., 2012) presented the first mtDNA data set which provides full

control region sequences from the Ghanaian population. This is of great forensic significance due to the real underrepresentation of Africa mtDNA data in general. However, data on Ghanaian population autosomal STRs is lacking.

Autosomal STRs diversity can be used to identify individuals since it provides higher discriminating powers unlike than of the Y- STRs, X-STRs and the mtDNA which are linked to particular lineage groups (Ziętkiewicz *et.al.*, 2012). Ghana therefore need to equip itself well enough in the area of forensics since it is the back bone of criminal investigation in order to help identify perpetrators as well as crime victims and family relations among others. It is for this sole that the Ghana Police Forensic Lab was set up and commissioned by the European Union in December, 2011 (Ghana web).

Determination of the allele frequency of the Ashanti population of Ghana using some selected autosomal STR markers would be of greater interest. This in effect will help build a startup data in generating an allele frequency database for the various STR markers. The FGA and D18S51 are part of the most polymorphic STR markers among various studied populations (Butler, 2006) and would therefore be used in this study.

2.7 STUDY MARKERS

2.7.1 FGA

FGA, also called FIBRA or HUMFIBRA is a tetranucleotide compound repeat located on the fourth human chromosome. The locus has a CTTT repetitive motif flanking both sides by degenerate repeats. The number of alleles reported so far range between 12.2 to 51.2 repeats. The size of the fragments ranges between 154bp and 314bp. The x.2 is as a

result of a 2-bp deletion of a CT in the region just before the core repeat (Butler & Hill, 2012; Butler, 2011) as described by http://www.cybertory.org/resources/CODIS/index.htmlm and http://www.cstl.nist.gov/biotech/strbase/fbicore.htm.The FGAhas the forward (F) and reverse (R) primer pairs as F: 5'-GCCCCATAGGTTTTGAACTCA-3' and R: 5'-TGATTTGTCTGTAATTGCCAGC-3' respectively.

2.7.2 D18S51

D18S51is a simple repeat made up of tetranucleotide repeat motif of AGAA. It is located on the eighteenth (18th) human chromosome. The D18S51 alleles range from 5.3 to 40 repeats and or 7 to 27 repeats. Its forward and reverse primer sequence are F: 5'-CAA ACC CGA CTA CCA GCA AC-3' and R: 5'-GAG CCA TGT TCA TGC CAC TG-3' with a base length of 234 and its fragment size ranging between 262-342bp.(http://www.cybertory.org/resources/CODIS/index.htmlmand http://www.cstl.nist.gov/biotech/strbase/fbicore.htm).This locus has been reported to have shown more alleles beyond 70 and hence making it very polymorphic. It also has a x.2-bp deletion due to a loss of AG at the 3' end (Butler & Hill, 2012; Butler, 2011).

2.8 The Asante Population of Ghana

The Ashanti (Asante) population of Ghana represents the largest sub groups of the Akan ethnic group of Ghana with a percentage of 30.1%. They comprise about 14.8% of the Ghanaian community by birth and are mostly located in the Ashanti Region of Ghana which is found in the central belt of the country between longitudes 0.15W and 2.25W,

and latitude 5.50N and 7.46N (http://www.ghana.gov.gh/index.php/about-ghana/regions/ashanti).

History has it that the Asante people of Ghana migrated through Bono Manso currently Techiman (Brong Ahafo Region) to the present-day Ashanti Region and had stretched beyond various boundaries through wars and the capture of land from other ethnic groups to achieve the current boarders of the Ashanti Region. Due to these wars, they have expanded their territories and are now located in numerous regions of the country such as Brong Ahafo and the Eastern Regions of Ghana (Nyanteng *et.al.*, 2013).

The Asante people instigates from eight clans namely Oyoko, Bretuo, Aduana, Asona Asene, Agona, Asakyi and Ekuona. Through unity, the Ashanti people grew up to be very powerful during the era of the first King (Asantehene), Nana Osei Tutu 1with the help of his friend Okomfo Anokye believed to have commanded the Golden Stool to descend from the sky. This stool symbolizes the unity and power of the Ashanti nation (Kumah 2009; Nyanteng *et.al.*, 2013). The people have their main language spoken to be Asante Twi (McCaskie, 2003).

CHAPTER THREE

MATERIALS AND METHODS

3.1 Study Site and Sampling

The study was carried among 60 unrelated Asante individuals in Kumasi in the Ashanti region of Ghana. The study site was the KNUST hospital. Permission was sought from the hospital administration to access the laboratory and patients coming to the laboratory. Asante patients who visited the laboratory were sampled randomly to constitute the study population.

A brief explanation of the study was made to these patients. Consent was sought from the selected patients and only those who consented were included in the study. Questionnaires were administered to the subjects and information on the individuals was collected. The questionnaire included basic information about the individual such as personal data of the recruit, the parents, grandparents, regions, districts and hometowns (Appendix). Also inquiries of the person having a unique family name. The unique family name here meant that the person's family had a particular name which identifies all members of that family.

Persons with both parents being Asantes or at least mother and grandmother being an Asante were recruited due to the high rate of ethnic and tribal intermarriage systems in Ghana. In all, Sixty (60) individuals consented to the study and had their blood samples (blood) taken.

3.2 Sample Collection and Storage

Two milliliters (2 ml) of intravenous blood were collected from each subject using a hypodermal needle syringe and stored in ethylene diamine tetra acetate (EDTA) vacutainer tubes and labeled appropriately. To each 2 mL of the blood, an equal volume of 4 M urea was added. This was to preserve the blood samples from being degraded. The blood was later stored in the fridge at -20^{o}C at the Molecular Biology Laboratory of the Department of Biochemistry and Biotechnology, KNUST until DNA isolation was done.

3.3 Genomic DNA Isolation

Prior to the extraction of the genomic DNA from each of the blood samples, 2 mL Eppendorf tubes were numbered 1 to 60 to correspond to the code indicated on the EDTA tube containing the blood sample. These were arranged on racks and 200 µL of the blood were transferred into tubes. DNA was extracted from each of the 60 blood samples using the magnetic beads method. To extract the DNA from each blood sample, 200 µL of blood was picked into the eppendorf tubes, 200 µL of lysis buffer BLM (LC Genomics) and 20 µL of protease K were added. This was then mixed thoroughly with a pipette set to 350µL by pipetting up and down five times.

The samples were then incubated at 55 ^{o}C for 10 mins using water bath and then allowed to cool to room temperature. 200 µL of ethanol was then added to each sample and 20 µL of the magnetic beads suspension added. The pipette was set to 550 µL and pipetted up and down five times to mix the resulting solution. To allow for sufficient binding of

the DNA to the magnetic beads, the solution was incubated for 2 mins at room temperature and the sample tubes brought in contact with a magnet for 1 min to allow the magnetic particles form pellets with the DNA. The supernatant was removed and discarded leaving only the pellets in the tube.

The pellets were resuspended in 720 µL of wash buffer BLM 1. These were then incubated at 55 ^{o}C for 10 mins using water bath and vortexed periodically to agitate the samples. The tubes were brought in contact with a magnet again for 1 min at room temperature to allow the formation of pellets. The supernatant was removed and discarded and 720 µL of wash buffer BLM 2 added to re-suspend the pellet and mixed using pipettes set to 650µL. These were then incubated at 55 ^{o}C for 10 mins again using water bath and vortexed periodically to agitate the samples. The tubes were brought in contact with a magnet for 1 min at room temperature to allow the formation of pellets. Wash buffer BLM 2 was added and after removing the supernatant, the pellets were dried at 55 ^{o}C for 10 mins using the water bath with the sample tubes left open to allow evaporation of all solutions present.

The dry pellets were re-suspended in 200 µL of elution buffer BLM and mixed thoroughly by setting the pipette to 150 µL to mix up and down for five times. These were then incubated at 55 ^{o}C for 10 mins using the water bath and agitated periodically using a vortex. The samples tubes were again brought into contact with magnet at room temperature for 3 mins to allow the particles form a pellet. 180 µL of the elute (containing the DNA extract) was then removed and transferred into a new sample tube.

The Eppendorf tubes were also labeled 1 to 60 respectively and stored at -20 ^{o}C. Prior to quantification and amplification, all extract were visualized on 0.7% agarose gel stained with ethidium bromide to confirm the presence and quality of DNA.

3.4 DNA Quantification

The concentrations of the extracted DNA were quantified using Nano Drop 1000 spectrophotometer. This equipment uses an operating software package (ND-1000 version 3.0 Thermo Fisher Scientific Inc., Waltham, MA, USA).

Blank was read by pipetting 1µL-2µL of the elution buffer used in extracting the DNA from the blood samples. The lens was cleaned after every 15-20 measurements to recalibrate the Nano Drop. The frozen DNA samples were thawed, vortexed and spun briefly to remove drops from the walls and lid of the tubes. The concentrations and purity of the samples were then measured. Firstly, they were mixed by gently flicking the base of the sample tubes containing the DNA extracts about four (4) to five (5) times. The pipette was set to 2µL and this amount aliquoted onto the pedestal/lower arm.

The ID number for each sample was then entered onto the computer software once the sample is aliquoted onto the pedestal and closed. The sample type (DNA-50'dsDNA) was selected for each sample after which the samples were measured by clicking 'measure' to record the concentration and purity of the DNA into a spreadsheet.

For each sample, the concentration reading was done three (3) times and the average calculated. After every fifteen (15) samples, the re-blanking with the elution buffer was

done to avoid contamination. The lenses were cleaned after all samples had been run by aliquotting 2-3 μL of water onto the top and lower arms and subsequently blotting with tissue wipe.

3.5 PCR OPTIMIZATION AND VISUALIZATION OF AMPLICONS ON AGAROSE GEL

3.5.1 PCR Optimization of Annealing Temperatures for each Study Marker

The DNA markers used for this study were FGA and D18S51 with the primer sequences as described on (http://www.cybertory.org/resources/CODIS/index.htmlm and http://www.cstl.nist.gov/biotech/strbase/fbicore.htm) with expected band sizes of 158-314 and 262-342 base pairs for FGA and D18S51 respectively. The FGA forward primer F: 5'-GCCCCATAGGTTTTGAACTCA-3' (Tm = 55.2 ^{o}C) and reverse primer R: 5'-TGATTTGTCTGTAATTGCCAGC-3' (Tm = 53.8 ^{o}C) were used together with the D18S51 forward primer F: 5'-CAA ACC CGA CTA CCA GCA AC-3' (Tm = 56.2 ^{o}C) and reverse primer R: 5'-GAG CCA TGT TCA TGC CAC TG-3' (Tm = 56.7 ^{o}C).

To optimize the amplification process, gradient PCR was done using varying annealing temperatures of 56 ^{o}C, 58 ^{o}C, 60 ^{o}C and 62 ^{o}C for each primer set. The PCR reaction mix was carried out using the same DNA sample but different primer sets for each annealing temperature to be run. For each 0.5 mL PCR tube containing a 25 μL reaction volume, there was 2 μL of template DNA, 2.5 μL of PCR buffer, 0.5 μL of dNTP, 0.5 μL of each

forward and reverse primer of the FGA marker, 0.25 μL of Taq polymerase and 18.75 μL of deionized water. This was then repeated for D18S51 marker using the same volumes. For each marker, four reaction volumes were prepared for each of the annealing temperatures of 56 ^{o}C, 58 ^{o}C, 60 ^{o}C and 62 ^{o}C. The DNA amplification was carried out in an MJ Research PTC-200 thermal cycler. It was programed to run gradient PCR such columns 5, 6, 8 and 12 contained samples to be annealed at 56 ^{o}C, 58 ^{o}C, 60 ^{o}C and 62 ^{o}C, respectively with their initial denaturation for five (5) minutes at 95 ^{o}C. 30 seconds cyclic denaturation at 95 ^{o}C, 30 seconds primer annealing at temperatures 56 ^{o}C, 58 ^{o}C, 60 ^{o}C and 62 ^{o}C and a 30 seconds extension stage at 72 ^{o}C. These three stages were repeated for 30 consecutive cycles after which a final extension of 5mins at 72 ^{o}C. The products were then held at 4 ^{o}C until gel electrophoresis was carried out. These were then run on 2.5% agarose gel with DNA of known band sizes as control samples.

3.5.2 Duplex PCR Optimization

Gradient PCR amplification was carried out again on the same DNA used for the initial amplification and subjected to the same conditions of the initial amplification but now all two markers (four primer sets) were combined in a reaction mix and run at annealing temperatures of 58 ^{o}C and 60 ^{o}C only. The 25 μL reaction volume was prepared such that each 0.5 mL PCR tube contained 2 μL of template DNA, 2.5 μL of PCR buffer, 0.5 μL of dNTP, 0.5 μL of each forward primer for the two markers, 0.5 μL of each reverse primer for the three markers, 0.3 μL of Taq polymerase and 17.70 μL of deionized water. The products were run on 2.5% agarose and visualized using UV transilluminator.

3.6 PCR Amplification

For all sixty (60) DNA extracted and quantified samples, PCR amplification was carried out increasing the reaction volume from 25 μL to 40 μL. To each 0.5 mL PCR tube, a reaction mix containing 40 μL reaction of each sample was prepared for amplification. This included 2 ng of template DNA, 4 μL of PCR buffer, 0.8 μL of dNTP, 0.8 μL of each forward primer, 0.8 μL of each reverse primer, 0.4 μL of Taq polymerase and 29.6 μLof deionized water. The DNA amplification was carried out. This time, it was programed such that the initial denaturation for five (5) mins was at 95 ^{o}C. 30 seconds cyclic denaturation at 95 ^{o}C, 30 seconds primer annealing 60 ^{o}C and a 30 seconds extension stage at 72 ^{o}C. These three stages were repeated for 30 consecutive cycles after which a final extension of 5mins at 72 ^{o}C. The products were then held at 4 ^{o}C and later transferred into -20 ^{o}C freezer to prevent destruction of the DNA amplicons until gel electrophoresis was carried out.

3.7 GEL AND STAIN PREPARATIONS

3.7.1 Apparatus Setup

Polyacrylamide gel electrophoresis was run on the Mini-Protean ®Tetra cell electrophoresis system (Bio-Rad, US). Before the preparation of the polyacrylamide gel, the gel cassette was set following the following steps. The gel glass plates, plastic plate (for balancing), grey casting stand and the casting frame were cleaned and dried with tissue wipe. The short glass plate was slide along the spacer plate (forming the gel cassette sandwich) ensuring that the label on the spacer plate was up and the short plate

facing the casting frame. The gel cassette sandwich was placed on the casting gasket with the grey casting stand beneath it to seal and prevent leakage of the gel when poured. To ensure that there was no leakage, the gel cassette frame was filled with distilled water to the brim of the short plate. The water was then poured out and the excess distilled and dried using filter paper. The glass cassette sandwich was then filled with the 5% polyacrylamide gel prepared (as described in **Section 3.7.2** below). The comb was then set in place to create wells for loading of the samples. This was left on the working bench for 30mins to allow the gel polymerize.

3.7.2 Preparation and Casting of Polyacrylamide Gel

Polyacrylamidegel (5%) was prepared using chemicals and reagents from Thermo Scientific Corporation. A total volume of 10 mL 5% polyacrylamide gel was prepared and cast for the visualization of PCR products. To prepare this, 1000 µL of 10X Tris Borate EDTA buffer (10X TBE) was pipetted into a 50 mL tube. 2500 µL of 20 % Acrylamide / Bisacrylamide and 6500 µL of distilled water was added. This was vigorously agitated using a magnetic stirrer for 1min to ensure complete mixing. 12 µL of TEMED was added and the tube swirled several times for thorough mixing. 60 µL of 10% Ammonium Persulfate (APS) was added and mixed immediately since polymerization starts right after the addition of the APS. The mixture was then carefully pipetted or poured in between the glass gel plate spacers making sure to avoid the formation of bubbles and a comb inserted to create wells for loading of the PCR products. This was left to polymerize for 30mins. The gel plate spacers was cleaned to take away any excess acrylamide.

3.7.3 Preparation of Ethidium Bromide for Staining of the Gel

0.5 $\mu g.mL^{-1}$ of aqueous ethidium bromide solution was prepared by adding 3 µL of 10 $mg.\mu L^{-1}$ of ethidium bromide stock solution into 50 mL of 1XTBE buffer. After the 105 mins, the power supply for the electrophoresis was stopped. The polyacrylamide gel was carefully removed from the glass plates and placed in the 0.5 $\mu g.mL^{-1}$ of the ethidium bromide stain prepared. The gel was then stained for about 15-30 mins in the Ethidium bromide stain and washed in 1X TBE for 15 mins before visualizing it on the UV transilluminator.

3.8 OPTIMIZATION OF POLYACRYLAMIDE GEL FOR ELECTROPHORESIS

The 5% PAGE was optimized for several times to identify the best time at which the PCR products could separate well on the gel. The gel was prepared, cast and run as stated in **sections 3.7.2**. This was warmed and pre-run at 100V for 30mins after which the five samples together with a 1kb full scale DNA ladder were loaded and run at 100V initially for one hour (1hr) and then subsequently for $1^{1}/_{2}$ hrs and 2 hrs. An ice pack was placed in the electrophoretic buffer to prevent over heating of the gel. After electrophoresis, the gel was stained for 15 mins in ethidium bromide solution and visualized using UV transilluminator.

3.9 Polyacrylamide Gel Electrophoresis

After the various gel optimizations, good separation and much sharper bands were resolved from the PAGE run at 100V for 120 mins. All sixty (60) PCR products were therefore run alongside 1 kb full scale ladder on a 5 % polyacrylamide gel at 100V for 120 mins. Afterwards, they were stained in ethidium bromide solution for 15 mins and then images of the gels visualized and captured using the UV transillumnator.

3.10 Statistical Analysis and Generation of Allele Frequency Distribution

Upon the PAGE analysis, the alleles and their respective number of base pairs for the study markers, FGA and D18S51 were generated and calculated manually using the 1 kb DNA ladder as a guide. The different alleles were identified by measuring the distance to which the bands had travelled from the wells of the gel and calculating their number of base pairs proportionally by comparing to the distance travelled by the bands of the ladder. The number of base pairs was then used in generating the alleles making reference to http://www.cstl.nist.gov/biotech/strbase/fbicore.htm. STR allelic frequency was then tabulated in a Microsoft excel sheet and the various frequencies of the alleles was calculated. The heterozygosity ratio, expected heterozygosity ratio and the polymorphism information content for both FGA and D18S51 markers were calculated.

CHAPTER FOUR

RESULTS

4.1 Study Population

The individuals selected for the study comprised of Asantes from various Ashanti towns and districts in the Ashanti region. The study population had both their parents being Asantes as well as their grandparents. This was to have individuals who pure Asante's background. A few had their fathers and grandfathers hailing from other Akan tribes such as Fante and Akuapim.Of the sixty (60) individuals, thirty five (35) were females and twenty five (25) were males with their ages ranging between eighteen (18) and sixty five (65) years. Five (5) persons had unique family names.

4.2 Genomic DNA isolation

DNA extraction was carried out using the BLM magnetic bead protocol as detailed in **Section 3.3** of this thesis. The extracted DNA was suspended in 200 µL of the BLM elution buffer and 180 µL of it aliquoted for storage at -20 ^{o}C for subsequent analysis. The DNA extracts were visualized on 0.7% agarose gel electrophoresis to confirm the extraction of DNA at the end of the extraction process. The agarose gel electrophoresis gave a positive feedback on the extraction with distinct bands (image not shown).

4.4 DNA quantification

After the quantification of the extracted DNA, most of the concentration of the DNA extracted ranged between 2 $ng.\mu L^{-1}$ and 15 $ng.\mu L^{-1}$with a few measuring above 20 $ng.\mu L^{-1}$. Calculating the purity from the absorbance ratio, about half the DNA samples had their purity ranging between 1.8 and 2.0 with the rest being below 1.8. The results of the quantification are tabulated in **Table 4.1**.

Table 4.1 DNA Quantification

SAMPLE ID	CONC. ($ng.\mu L^{-1.}$)	Absorbance (A260/280)	SAMPLE ID	CONC. ($ng.\mu L^{-1.}$)	Absorbance (A260/280)
1	30.16	1.61	31	9.37	1.71
2	6.79	1.28	32	10.64	1.28
3	2.52	1.61	33	6.94	1.09
4	5.9	1.71	34	14.39	1.73
5	3.32	2	35	27.52	1.59
6	2.72	2.04	36	6.76	1.63
7	5.66	1.2	37	12.79	1.9
8	2.38	1.68	38	4.31	1.3
9	2.58	2.08	39	4.18	1.92
10	8.2	1.65	40	59.02	1.06
11	3.63	1.43	41	2.4	2.03
12	3.04	1.54	42	15.83	1.38
13	3.12	2.09	43	5.35	1.65
14	42.46	1.66	44	10.99	1.28
15	3.31	1.86	45	12.11	1.78
16	3.1	1.28	46	15.47	1.54
17	2.12	2.94	47	9.46	1.92
18	2.61	2.68	48	15.92	2.38
19	1.58	0.82	49	8.98	1.47
20	1.35	1.96	50	50.78	3.32
21	19.79	2.16	51	25.98	1.68
22	7.04	2.11	52	18.37	1.49
23	6.34	1.81	53	28.46	1.52
24	1.74	1.84	54	2.89	1.78
25	1.58	1.63	55	28.32	1.68
26	3.89	1.93	56	22.77	1.64
27	42.99	0.62	57	14.85	1.48
28	1.62	1.72	58	20.15	1.43
29	0.85	1.9	59	34.79	1.41
30	14.93	1.21	60	3.35	1.96

Table showing the sample identification (Sample ID), the concentration of DNA in $ng.\mu L^{-1}$and the purity (Absorbance, A260/A280) of the DNA samples with most of the concentrations ranging between 2 $ng.\mu L^{-1}$and 30 $ng.\mu L^{-1}$ and about half of them having their purity ranging between 1.8 and 2.0.

4.5 PCR OPTIMIZATION AND VISUALIZATION OF AMPLICONS ON AGAROSE GEL

4.5.1 PCR Optimization: Annealing Temperatures for each Study Marker

Optimization of PCR was done to know the appropriate PCR conditions such as annealing temperature favorable to each of the markers FGA and D18S51 primer pairs being used for the study. For each marker primer pair, gradient PCR was run at annealing temperatures of 56 ^{o} C, 58^{o}C, 60^{o}C and 62^{o}C. Bands were present for all the temperatures indicating that amplification was successful using all the temperatures run. As shown in **Plate 4.2,** the bands were visible for the markers D18S51 and FGA. The bands represent PCR products annealed at temperature 56 ^{o}C, 58 ^{o}C, 60 ^{o}C and 62 ^{o}C respectively.

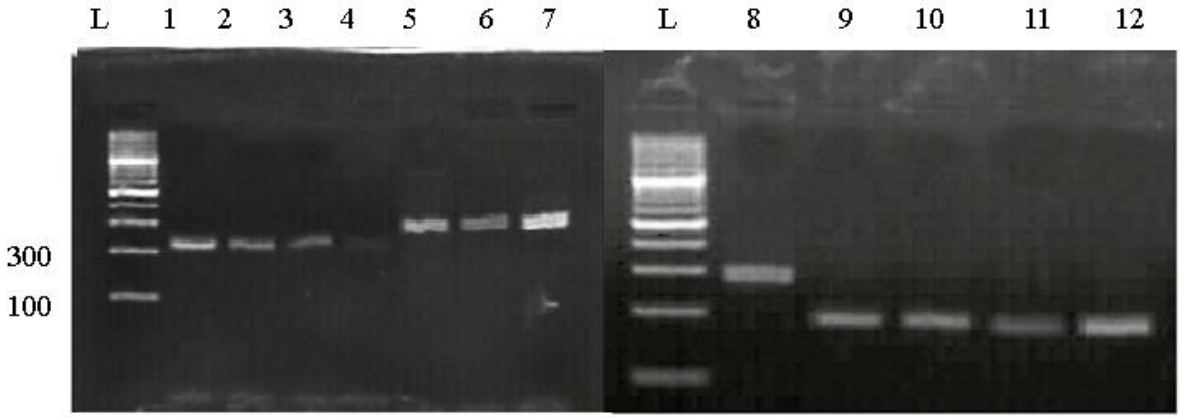


Plate 4.1 2.5% Agarose Gel Electrophoresis Profile of PCR Product with Primers at Different Annealing Temperatures

A DNA sample (26) of 2 ng was used as a template in a PCR reaction with primers FGA and D18S51 at annealing temperatures 56 ^{o}C (lanes 1, 5 and 9), 58 ^{o}C (lanes 2, 6 and 10), 60 ^{o}C (lanes 3, 7 and 11) and 62 ^{o}C (lanes 4, 8 and 12). The PCR product (5 μL) of each sample was loaded alongside with 100 bp ladder on 2.5% agarose gel electrophoresis. The electrophoresis results showing bands for 100 bp ladder (L), a control sample and PCR products for both study markers (D18S51 and FGA) at various annealing temperatures of 56 ^{o}C, 58 ^{o}C, 60 ^{o}C and 62 ^{o}C . The control sample was

loaded onto lanes 1, 2, 3 and 4, D18S51 onto lanes 5, 6, 7 and 8 and FGA onto lanes 9, 10, 11 and 12, respectively.

4.5.2 PCR Optimization: Duplex PCR of the Two Markers, FGA and D18S51

In choosing the best annealing temperature for D18S51 and FGA put together, the sharpness (brightness) of the bands produced when run separately were considered. The amplification was therefore repeated, for temperatures at 58 ^{o}C and 60 ^{o}C since their bands appeared sharper compared to the rest, hence the decision to choose the two temperatures. The PCR were therefore repeated again but this time at annealing temperatures of 58 ^{o}C and 60 ^{o}C. The sample contained both two markers for the study, FGA and D18S51. The PCR products were run on 2.5% agarose gel.

Having achieved good amplification which produced distinct visible bands on agarose electrophoresis for all the markers, duplex PCR was run for both the two markers at an annealing temperature of 58 ^{o}C. The results for this were not satisfactory when analysed on agarose gel electrophoresis. The duplex PCR reaction was repeated at an annealing temperature of 60 ^{o}C which produced distinct bands on agarose gel electrophoresis (**Plate 4.3**)**.** The annealing temperature of 60 ^{o}C was chosen as best for duplex PCR for the two markers combined.

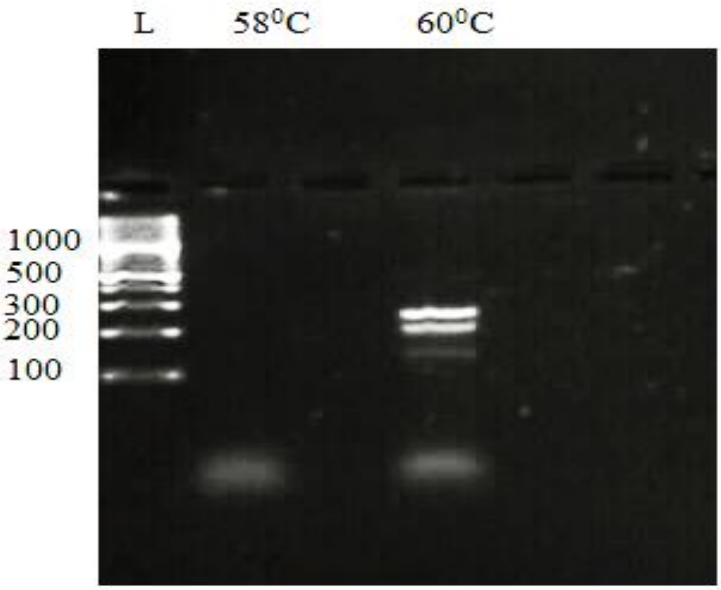
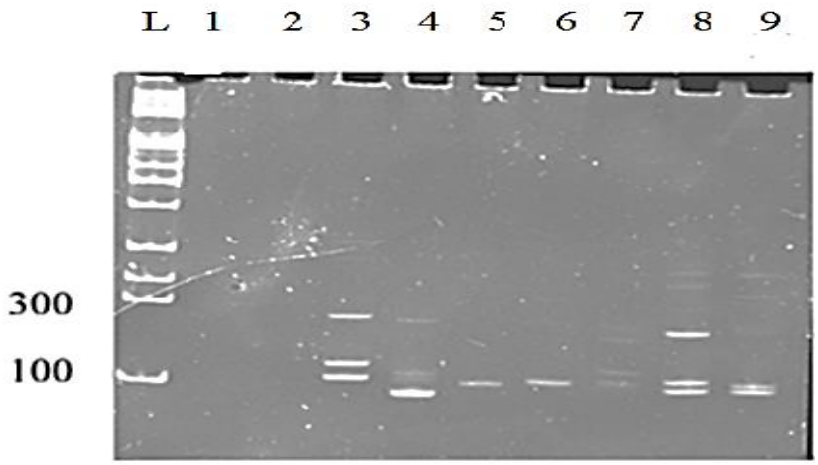


Plate 4.2 Electrophoresis of PCR Product Annealed at 60°C

Image of PCR products of both set of STR markers amplified at annealing temperatures of 58 °C and 60 °C and analysed on 2 % agarose gel together with a 100 bp ladder on lane L. No band was visible for the duplex with annealing temperature at 58 °C.

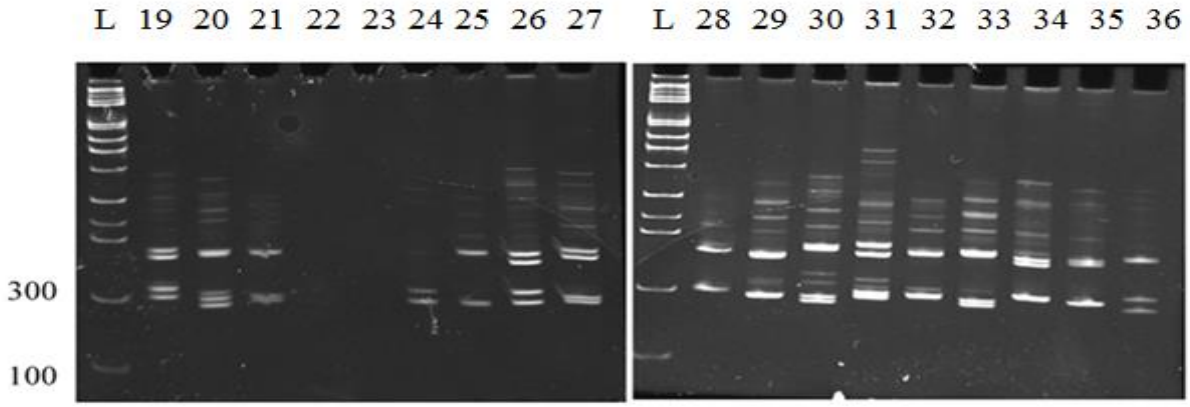
4.6 Polyacrylamide Gel Electrophoresis.

Upon the various PAGE optimizations, it was concluded that the samples be run on 5% polyacrylamide gel at a voltage of 100 volts for 120 min. All the PCR products were then subjected to these conditions. All 60 PCR product samples analysed showed visible bands with some being very sharp. The results are as shown in **Plates 4.3a, b, c and d.**



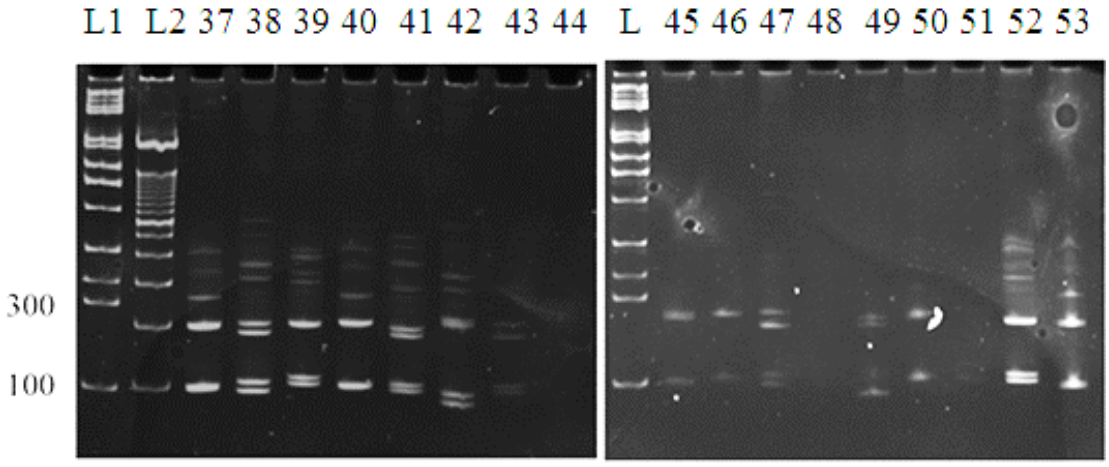
Plates 4.3a Polyacrylamide Gel Electrophoresis

Diagrams showing bands for ladder (L) and PCR products for samples (1to 9). These were run on 5% polyacrylamide gels at 100V for 120 mins along with 1kb full scale DNA ladder.



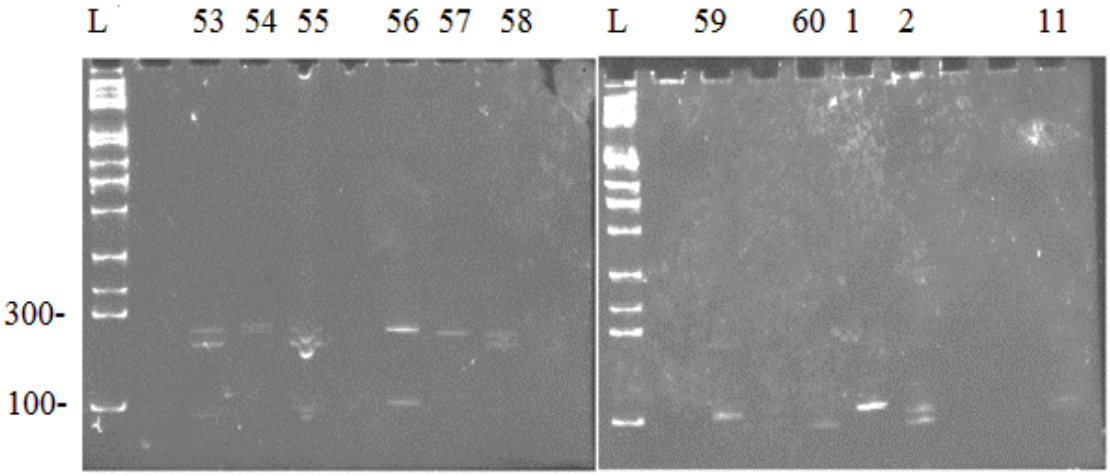
Plates 4.3b Polyacrylamide Gel Electrophoresis

Diagrams showing bands for ladder (L) and PCR products for samples (19 to 36). These were run on 5% polyacrylamide gels at 100V for 120 mins along with 1kb full scale DNA ladder.



Plates 4.3c Polyacrylamide Gel Electrophoresis

Diagrams showing bands for ladder (L) and PCR products for samples (37 to 53). These were run on 5% polyacrylamide gels at 100V for 120 mins along with 1kb full scale DNA ladder.



Plates 4.3d Polyacrylamide Gel Electrophoresis

Diagrams showing bands for ladder (L) and PCR products for samples (1, 2, 11 and 53 to 60). These were run on 5% polyacrylamide gels at 100V for 120 mins along with 1kb full scale DNA ladder.

4.7 Statistical analysis and allele frequency distribution

Of the 60 samples amplified and run using 5% polyacrylamide gel electrophoresis, all samples showed bands, suggesting alleles. The electrophoresis of PCR samples for the population allele frequency study of the FGA and D18S51 markers revealed a range of 1 to 4 alleles per sample analysed. Over all, a total of 23 allele types (14 for FGA and 9 for D18S51) were observed as 144 bands (alleles) for the two markers, with FGA having 79 alleles and D18S51 having 65 alleles. Most of the bands from the gel images were sharp indicating strong amplification with few being faint (weak amplification). The different alleles and STR allelic frequency for the study markers or loci are presented on the **Table 2** and **3** below, showing the number of alleles for the FGA and D18S51 markers with their respective frequencies and the overall frequencies of the alleles.

Table 4.2 Allele Frequency Distribution of FGA Loci with repeat motif [CTTT]

Alleles	No. of bp	No. of Alleles	Repeat Motif	Frequency
43.2	282	3	$[TTTC]_4TTTT$ TT $[CTTT]_8$ $[CTGT]_5$ $[CTTT]_{13}[CTTC]_4[CTTT]_3CTCC[TTCC]_4$	0.037974684
41.2	274	4		0.050632911
39.3	267	9		0.113924051
38	260	12		0.151898734
36.1	253	19		0.240506329
34.2	246	9	$[TTTC]_4TTTT$ TT $[CTTT]_{18}[CTTC]_3[CTTT]_3$ $CTCC[TTCC]_4$	0.113924051
33.3	243	1	$[TTTC]_4TTTT$ TT $[CTTT]_{17}[CTTC]_3$ $[CTTT]_3$ $CTCC$ $[TTCC]_{4.X}$	0.012658228
33	240	9	$[TTTC]_4TTTT$ TT $[CTTT]_{16}[CTTC]_3$ $[CTTT]_3$ $CTCC$ $[TTCC]_{4.X}$	0.113924051
31.2	234	2	$[TTTC]_4TTTT$ TT $[CTTT]_{15}[CTTC]_3$ $[CTTT]_3CTCC$ $[TTCC]_4$	0.025316456
30.1	229	1	$[TTTC]_3TTTT$ $TTCT[CTTT]_{16}$ CCTT $[CTTT]_5CTCC[TTCC]_{2.X}$	0.012658228
28.3	223	2	$[TTTC]_3TTTT$ TT $[CTTT]_{21}CTCC$ $[TTCC]_{2.X}$	0.025316456
27.2	218	4	$[TTTC]_3TTTT$ $TT[CTTT]_{20}CTCC[TTCC]_2$	0.050632911
26.1	213	3	$[TTTC]_3TTTT$ $TTCT[CTTT]_{18}CTCC$ $[TTCC]_{2.X}$	0.037974684
25.1	209	1	$[TTTC]_3TTTT$ $TTCT[CTTT]_{17}CTCC$ $[TTCC]_{2.X}$	0.012658228
Total number of alleles		79		

Table showing the alleles present for the FGA marker with its respective number of alleles, number of base pairs, repeat motif and frequency distributions. 'X' representing an additional nucleotide C or Tin repeat motif. Both the frequency of each allele at the locus and the frequency of the marker within the study population are indicated
*(*http://www.cybertory.org/resources/CODIS/index.htmlm and http://www.cstl.nist.gov/biotech/strbase/fbicore.htm)

Table 4.3 Allele Frequency Distribution of D18S51 Loci with repeat motif [AGAA]

Alleles	No. of bp	No. of Alleles	Repeat Motif	Frequency
30.2	356	7	[AGAA]30. XX	0.107692308
27.1	343	13	[AGAA]8.X	0.2
24.1	331	4	[AGAA]24.X	0.061538462
21.2	320	11	[AGAA]21.X X	0.169230769
19	310	13	$[AGAA]_{19}$	0.2
16.2	300	8	[AGAA]16.X X	0.123076923
14.1	291	4	[AGAA]14.X	0.061538462
10	274	4	$[AGAA]_{10}$	0.061538462
8.1	267	1	[AGAA]8.X	0.015384615
Total number of alleles		65		0.488721805

Table showing the alleles present for the D18S51 marker with its respective number of alleles, number of base pairs, repeat motif and frequency distributions.'X' representing the additional nucleotide A or G. Both the frequency of each allele at the locus and the frequency of the marker within the study population are indicated.

Based on the genetic and forensic efficiency parameters calculated from the allelic frequencies of the markers, the markers had their most polymorphic alleles being alleles 36.1 and (19 and 27.1) for FGA and D18S51 respectively. These alleles had frequencies of 0.240506329 and 0.2 for FGA and D18S51 respectively with allele 36.1 occurring 19 times and alleles 19 and 27.1 occurring 13 times as shown on **Table 2** and **3** respectively.

The number of heterozygote alleles for both FGA and D18S51 were 24 and 14 respectively with their observed heterozygosity (Ho) which is the proportion of the number of heterozygotes alleles to the total number of alleles for a particular loci or marker being 0.3037974684 and 0.2153846154 respectively.

The expected heterozygosity (He) was recorded as 0.870373337 and 0.853017751 with respect to FGA and D18S51. The homozygosity which is also the sum of squares of all allele frequencies at a locus was recorded as 0.129626663 and 0.146982249 for FGA and D18S51 correspondingly. Again the polymorphism information content (PIC) which is the measure of the informativeness of a genetic marker was also calculated as 0.85797208 and 0.83584121 for FGA and D18S51 markers respectively as shown in **Appendix 3** and **4**.

CHAPTER FIVE

DISCUSSION

This research was conducted with the objective of assessing the allele distribution of FGA and D18S51 STR markers within the Asante population of Ghana and also to provide information for the generation of a DNA database for the study population using the study markers. In this study, human blood samples were collected and DNA isolated, quantified, amplified and analyzed using 5% polyacrylamide gel electrophoresis.

The few individuals who had their fathers and or grandfathers hailing from other Akan tribes were included in the study since Asantes patronizes matrilineal linage. Also, the Akans on a whole are noted to be one people who were divided into sub tribes due to areas where they migrated to during wars (as cited in Nyanteng *et.al.*, 2013).

The isolation of genomic DNA using the BLM protocol in this study yielded good amount of DNA attesting to Butler (2012) which states that for any successful molecular maker analysis, the isolation of high genomic DNA is very essential, hence the quantity or amount of DNA to be used in PCR is important for downstream application in forensic investigation (Butler, 2011).

Majority of the DNA extracts that had their purity ratios ranging between 1.8 to 2.0 with a few being very low or high than the purity range (1.8-2.0) suggests low amounts of DNA extracts in these samples and or possible protein contaminations resulting from the

low effect of the activity of proteinase K on the extraction process of these samples. Upon the electrophoresis of the PCR products on the 5% polyacrylamide gel, all 60 PCR product samples used in this study showed various bands indicating the presence of the amplified sites of these individuals based on the FGA and D18S51 forensic markers used.

Although this study was in accordance with a study by White *et al.,* 1997 which stated that most STR markers analyzed as DNA fragments are made up of about 100-350 base pairs as seen in this study, there were few alleles which had their base pairs being below 100 bp as seen on **Plates 4.3a - 4.3d** of **Chapter 4**. This suggests that within the Asante population, there are alleles of FGA and or D18S51 having less than 100 bp which have not been reported in the STRBase as available onhttp://www.cstl.nist.gov/biotech/strbase/.

Of the few bands (alleles) observed as faint bands, their concentrations were noted to have their purity ratios going out of the accepted range (1.8 – 2.0) and being low. For instance, PCR products numbered 7, 57 and 58 which had their purity ratios to be 1.2, 1.48 and 1.43 respectively which explains why they showed very faint bands (alleles). This suggests a possibility that the faint bands could have resulted from the fact that the purity of the amount of DNA extract used in the PCR reaction was less.

The reduction in the number of bands (alleles) observed after electrophoresis could also have been as a result of these individuals being homozygote for these loci or could have resulted from some PCR issues such as null allele which are mostly encounted during

DNA STR analysis. For null allele to occur during PCR, the binding site which contains the region of interest which the primers bind to have other sequences which also contributes to the size of the amplified section of the DNA. In these regions of interest, variations or differences between the sequences either within or outside the repeat region where the primers bind may occur leading to missing alleles (Pompanot *et.al.,* 2005). In this study, an instance of null allele could result in the shortage of the number of bands or alleles (1, 2 or 3 bands) obtained for the FGA and D18S51markers.

Again in this study, the FGA and D18S51 markers showed higher diversity and polymorphism of alleles as in other studies incorporated into the STRBase. Comparing the forensic efficiency parameters in this study to that of a study population of the Mesan and Basra provinces in Southern Iraq which was also being studied for the first time by Hameed *et al.*, 2015, the homozygosity, heterozygosity and polymorphism information content (PIC) of these markers were almost the same. Within the Asante population, homozygosity, heterozygosity and PIC values of 0.1296, 0.8704 and 0.8580 were obtained for FGA loci whereas the study on these Southern Iraq populations had 0.141, 0.877 and 0.839 for FGA loci respectively. In the Southern Iraq populations, the total number of observed allele types for FGA was 8 and 11 for D18S51 (Hameed *et.al.,* 2015) where 13 and 11 different alleles types were observed for FGA and D18S51 respectively in the Asante population.

Study conducted in Botswana on five sub populations for the first time also reported 12 allele types for FGA and 17 for D18S51. The homozygosity, heterozygosity and polymorphism information content (PIC) for these Botswana populations were 0.141, 0.5859 and 0.84 respectively for FGA and 0.131, 0.839 and 0.84 for D18S51 respectively (Mpoloka *et.al.*, 2008). This was also in accordance with the Asante population of Ghana and hence making these markers to being highly polymorphic as in other studies.

Again, the number of allele types observed in this study of the Asante population was in agreement with a study by Kashyap *et.al.*,(2004), which stated that most STR markers generally have about seven (7) to thirty (30) different alleles on the average and are highly polymorphic in various populations (Butler, 2006).

CHAPTER SIX

CONCLUSION AND RECOMMENDATION

6.1 Conclusion

The general conclusions of this study are:

- Overall, 14 and 9 different alleles were determined for the FGA and D18S51 markers respectively.

- A frequency distribution of FGA and D18S51 STR markers among the Asante population of Ghana has been established with the FGA locus recording a frequency distribution of 0.54861111 and D18S51 recording 0.451388889.

- The findings of the study indicates that the FGA and D18S51 STR markers are important in forensic identification and can be used to estimate genotypic frequencies among the Asante population.

- Finally, the outcome of this study serves as a starting point for the generation of a database for STR markers among the Asante population.

6.2 Recommendations

- The frequency distribution of the FGA and D18S51 STR markers can be determined among other ethnic groups in Ghana.

- Mitochondrial DNA studies could be conducted on the individuals recruited as the study population to ascertain their true identity as Asante individuals.

- Additional STR markers can be studied among the Asante population.

- Also, further studies can be done with the FGA and D18S51 STR markers by employing the use of capillary electrophoresis.

- Again, other extraction methods can be used to enhance the quality and quantity of DNA extracted.

REFERENCES

Alexov, E. (2014). Advances in Human Biology : Combining Genetics and Molecular Biophysics to Pave the Way for Personalized.

Alkan, C., Coe, B. P., & Eichler, E. E. (2011). Genome structural variation discovery and genotyping. Nature Reviews Genetics, 12(5), 363–376.

Amatucci, J. (2010). Dna Fingerprinting. Worcester Polytechnic Institute.

Barbas, C. F., Burton, D. R., Scott, J. K., & Silverman, G. J. (2007). Quantitation of DNA and RNA. CSH Protocols, 2007(11), pdb.ip47. http://doi.org/10.1101/pdb.ip47

Barril, P., & Nates, S. (2012). Introduction to agarose and polyacrylamide gel electrophoresis matrices with respect to their detection sensitivities. Gel Electrophoresis - Principles and Basics, 3–15. http://doi.org/10.5772/38573

Bentayebi, K., Abada, F., Ihzmad, H., & Amzazi, S. (2014). Genetic ancestry of a Moroccan population as inferred from autosomal STRs. *Meta gene, 2*, 427-438.

Biesecker, L. G., Bailey-Wilson, J. E., Ballantyne, J., Baum, H., Bieber, F. R. & Brenner, C., (2005). EPIDEMIOLOGY: Enhanced: DNA Identifications After the 9/11 World Trade Center Attack. Science, 310(5751), 1122.

Blessmann, M., Kaifi, J. T., Schurr, P. G., Cihan, A., Kalinin, V. &Trump, F. (2008). Short tandem repeat polymorphism in exon 4 of esophageal cancer related gene 2 predicts relapse of oral squamous cell carcinoma. Oral Oncology, 44(2), 143–147.

Brown, P. O., & Botstein, D. (1999). Exploring the new world of the genome with DNA microarrays. Nature genetics, 21.

Budowle, B., Adamowicz, M., Aranda, X. G., Barna, C., Chakraborty, R. & Cheswick, D. (2005). Twelve short tandem repeat loci Y chromosome haplotypes: genetic analysis on populations residing in North America. Forensic Science International, 150(1), 1–15.

Budowle, B., Chakraborty, R., Giusti, A. M., Eisenberg, A. J., & Allen, R. C. (1991). Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE. American Journal of Human Genetics, 48(1), 137.

Budowle, B., Shea, B., Niezgoda, S., & Chakraborty, R. (2001). CODIS STR loci data from 41 sample populations. Journal of Forensic Sciences, 46(3), 453–489.

Butler, J. M. (2001). Forensic DNA Typing: Biology & Technology Behind STR Markers. Academic Press. Retrieved from https://books.google.com/books?hl=en&lr=&id=8sWJzjXc91wC&pgis=1

Butler, J. M. (2005). Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers. Academic Press. Retrieved from https://books.google.com/books?hl=en&lr=&id=gwDyBq2xLjIC&pgis=1

Butler, J. M. (2006). Genetics and genomics of core short tandem repeat loci used in human identity testing. Journal of Forensic Sciences, 51(2), 253–265. http://doi.org/10.1111/j.1556-4029.2006.00046.x

Butler, J. M. (2008). New resources for the forensic genetics community available on the NIST STRBase website. Forensic Science International: Genetics Supplement Series, 1(1), 97–99. http://doi.org/10.1016/j.fsigss.2007.10.035

Butler, J. M. (2009). Fundamentals of forensic DNA typing. Academic Press.

Butler, J. M. (2010). Fundamentals of Forensic DNA Typing. Fundamentals of Forensic DNA Typing, 363–396. http://doi.org/10.1016/B978-0-12-374999-4.00016-3

Butler, J. M. (2011). Advanced Topics in Forensic DNA Typing: Methodology: Methodology. Academic Press. Retrieved from https://books.google.com/books?hl=en&lr=&id=44tX9Ole490C&pgis=1

Butler, J. M. (2012). Advanced Topics in Forensic DNA Typing. Advanced Topics in Forensic DNA Typing, 271–292. http://doi.org/10.1016/B978-0-12-374513-2.00009-9

Butler, J. M. (2015). U.S. initiatives to strengthen forensic science & international standards in forensic DNA. Forensic Science International. Genetics, 18, 4–20. http://doi.org/10.1016/j.fsigen.2015.06.008

Butler, J. M., Buel, E., Crivellente, F., & McCord, B. R. (2004). Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. Electrophoresis, 25(10–11), 1397–1412. http://doi.org/10.1002/elps.200305822

Butler, J. M., Coble, M. D., & Vallone, P. M. (2007). STRs vs. SNPs: thoughts on the future of forensic DNA testing. Forensic Science, Medicine, and Pathology, 3(3), 200–205. http://doi.org/10.1007/s12024-007-0018-1

Butler, J. M., & Hill, C. R. (2012). Biology and genetics of new autosomal str loci useful for forensic DNA analysis. Forensic Science Review, 24(1), 16–26.

Calafell, F., Shuster, A., Speed, W. C., Kidd, J. R., & Kidd, K. K. (1998). Short tandem repeat polymorphism evolution in humans. European Journal of Human Genetics, 6(1), 38–49.

Chakraborty, R., & Kidd, K. K. (1991). The utility of DNA typing in forensic work. Science, 254(5039), 1735–1739.

Cortellini, V., & Cerri, N. (2011). Genetic variation at 5 new autosomal short tandem repeat markers (D10S1248 , D12S391) in a population- based sample from

Maghreb region, 368–371. http://doi.org/10.3325/cmj.2011.52.368

Crowley, L. (2011). Essentials of Human Disease. Jones & Bartlett Publishers. Retrieved from https://books.google.com/books?id=AH5Q8Sfg3_gC&pgis=1

Dale, W. M., Greenspan, O., & Orokos, D. (2006). DNA forensics: Expanding uses and information sharing. Search, National Consortium for Justice Information and Statistics.

Daniels, R., & Nicoll, L. (2011). Contemporary Medical-Surgical Nursing, Volume 1. Cengage Learning. Retrieved from https://books.google.com/books?id=-tmHtnxutW4C&pgis=1

David, P. C. (2005). Molecular Biology: Understanding the genetic revolution. Amsterdam: Elsevier Academic Press.

Diallo, K. (2013). Eyewitness Identification: how witnesses pick out a suspect? In Proceedings in GV-Global Virtual Conference.

Divne, A.-M., Edlund, H., & Allen, M. (2010). Forensic analysis of autosomal STR markers using Pyrosequencing. Forensic Science International: Genetics, 4(2), 122–129.

Donoghue, J. (2013). Police, Lies and Alibis: The True Story of a Front Line Officer. Troubador Publishing Ltd. Retrieved from https://books.google.com/books?id=9_ZJAgAAQBAJ&pgis=1

Edwards, A., Hammond, H. A., Jin, L., Caskey, C. T., & Chakraborty, R. (1992). Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics, 12(2), 241–253. http://doi.org/10.1016/0888-7543(92)90371-X

Edwards Jeanette, S. Carles.)2013(. European Kinship In The Age Of Biotechnology. Berghahn Books. Retrieved from https://books.google.com/books?id=BM5yYFj-xKgC&pgis=1

Egli, M., & Saenger, W. (2013). Principles of nucleic acid structure. Springer Science & Business Media. http://doi.org/10.1007/978-1-4612-51903

Ellegren, H. (2004). Microsatellites: simple sequences with complex evolution. Nature Reviews Genetics, 5(6), 435–445.

Erber, W. N. (Ed.). (2010). Diagnostic Techniques in Hematological Malignancies (Vol. 11). Cambridge University Press. Retrieved from https://books.google.com/books?id=4kyLvR96I3MC&pgis=1

Fendt, L., Röck, A., Zimmermann, B., Bodner, M., Thye, T., Tschentscher, F. & Parson, W. (2012). MtDNA diversity of Ghana: A forensic and phylogeographic view. Forensic Science International: Genetics, 6, 244–249. http://doi.org/10.1016/j.fsigen.2011.05.011

Ferrier, D. R. (2013). Biochemistry. Lippincott Williams & Wilkins. Retrieved from https://books.google.com/books?id=sSyMAwAAQBAJ&pgis=1

Gall John &Jason P.-J.)Ed.(.)2011(. Current Practice in Forensic Medicine. John Wiley & Sons. Retrieved from https://books.google.com/books?id=gkZdETOi7PwC&pgis=1

Gardner, T., & Anderson, T. (2015). Criminal Evidence: Principles and Cases. Cengage Learning. Retrieved from https://books.google.com/books?id=Ld3wBQAAQBAJ&pgis=1

Goodwin, W., Linacre, A., & Hadi, S. (2007). An Introduction to Forensic Genetics (Vol. 27). John Wiley & Sons. Retrieved from https://books.google.com/books?id=7wnivFnxel8C&pgis=1

Goodwin, W., Linacre, A., & Hadi, S. (2011). An Introduction to Forensic Genetics. John Wiley & Sons. Retrieved from https://books.google.com/books?id=OVxe_dp04x8C&pgis=1

Guilliatt, A. M. (2002). Agarose and polyacrylamide gel electrophoresis. In PCR Mutation Detection Protocols (pp. 1–12). Springer.

Hameed, I. H., Ommer, A. J., Murad, A. F., & Mohammed, G. J. (2015). Allele frequency data of 21 autosomal Short Tandem Repeat loci in Mesan and Basra provinces in south of Iraq. Egyptian Journal of Forensic Sciences. http://doi.org/10.1016/j.ejfs.2014.10.003

Hares, D. R. (2012). Expanding the CODIS core loci in the United States. Forensic Science International: Genetics, 6(1), e52--e54.

Hart, S. (2002). Using DNA to Solve Cold Cases. National Institute of Justice - Special Report, 1–32. Retrieved from papers2://publication/uuid/B09B9697-FAB9-42CF-8202-F65C0B57DCBC

Harty, L. C., Garcia-Closas, M., Rothman, N., Reid, Y. A., Tucker, M. A., & Hartge, P. (2000). Collection of buccal cell DNA using treated cards. Cancer Epidemiology Biomarkers & Prevention, 9(5), 501–506.

Hearn, R. P., & Arblaster, K. E. (2010). DNA extraction techniques for use in education. Biochemistry and Molecular Biology Education : A Bimonthly Publication of the International Union of Biochemistry and Molecular Biology, 38(3), 161–6. http://doi.org/10.1002/bmb.20351

Hosler, G. A., & Murphy, K. M. (2014). Molecular Diagnostics for Dermatology: Practical Applications of Molecular Testing for the Diagnosis and Management of the Dermatology Patient. Springer Science & Business. Retrieved from https://books.google.com/books?id=zA27BAAAQBAJ&pgis=1

http://www.cybertory.org/resources/CODIS/index.htmlm

http://www.cstl.nist.gov/biotech/strbase/fbicore.htm

http://alfred.med.yale.edu/alfred/flyer/ALFREDFlyer.pdf

(http://alfred.med.yale.edu/

http://www.hindawi.com/journals/ah/2014/924030/fig1/- assessed on 11/12/2015;3:47pm

(http://www.ghana.gov.gh/index.php/about-ghana/regions/ashanti; accessed on 21/12/15; 1:37pm

James, N. (2012). DNA Testing in Criminal Justice : Background , Current Law , Grants , and Issues.

Jeffreys, A. J., Brookfield, J. F. Y., & Semeonoff, R. (1985). Positive identification of an immigration test-case using human DNA fingerprints. Nature, 317(6040), 818–819.

Jeffreys, A. J., Wilson, V. & Thein, S. L. (1985). Hypervariable"minisatellite" regions in human DNA. Nature, 314(6006), 67–73.

Jiang, Y., Zhang, R., Lv, H., Li, J., Wang, M., Chang, Y. &Liu, G. (2013). HGPGD: The Human Gene Population Genetic Difference Database. PLoS ONE, 8(5), 1–6. http://doi.org/10.1371/journal.pone.0064150

Jobling, M. a., & Gill, P. (2004). Encoded evidence: DNA in forensic analysis. Nature Reviews Genetics, 5(10), 739–751. http://doi.org/10.1038/nrg1455

Jobling, M., Hurles, M., & Tyler-Smith, C. (2013). Human Evolutionary Genetics: Origins, Peoples & Disease. Garland Science. Retrieved from https://books.google.com/books?hl=en&lr=&id=5CwWBAAAQBAJ&pgis=1

Jutta Jutta. (2013). DNA Sequencing & The Human Genome Project. An Endeavor Revolutionizing Modern Biology. pdf. Retrieved November 16, 2015, from http://www.lehigh.edu/~inbios21/PDF/Fall2013/Jutta_11132013.pdf

Kashyap, V. K., Sitalaximi, T., Chattopadhyay, P., & Trivedi, R. (2004). DNA Profiling Technologies in Forensic Analysis, 4(1), 11–30.

Khodjet-el-Khil, H., Fadhlaoui-Zid, K., Gusmão, L., Alves, C., Benammar-Elgaaied, A., & Amorim, A. (2012). Allele frequencies for 15 autosomal STR markers in the Libyan population. Annals of Human Biology, 39(1), 80–83. http://doi.org/10.3109/03014460.2011.630678

King, R. C., Mulligan, P., & Stansfield, W. (2013). A dictionary of genetics. Oxford University Press.

Kirby, L. T. (1992). DNA fingerprinting. WH Freeman.

Kobilinsky, L. (Ed.). (2011). Forensic Chemistry Handbook (Vol. 17). John Wiley & Sons. Retrieved from https://books.google.com/books?id=eVsUB77_awMC&pgis=1

Krimsky, S., & Simoncelli, T. (2013). Genetic Justice: DNA Data Banks, Criminal Investigations, and Civil Liberties. Columbia University Press. Retrieved from https://books.google.com/books?id=kGDxQxyVRmQC&pgis=1

Kumah, D. (2009). Stools in Asante culture. school of graduate studies, Kwame Nkrumah University of science and Technology.

Lach, C. (2006). Dna fingerprinting.

Lewis, C. M. (2010). Hierarchical modeling of genome-wide Short Tandem Repeat (STR) markers infers native American prehistory. American Journal of Physical Anthropology, 141(2), 281–289.

Li, R. (2008). Forensic Biology: Identification and DNA Analysis of Biological Evidence. CRC Press. Retrieved from https://books.google.com/books?id=W1vMBQAAQBAJ&pgis=1

Linacre, A. M. T., Phillips, P., Wally, A., & Goodwin, W. (2001). STR data for the SGM Plus™ loci from three Indonesian populations. Forensic Science International, 123(2), 232–234.

Linacre, A., & Tobe, S. (2013). Wildlife DNA Analysis: Applications in Forensic Science. John Wiley & Sons. Retrieved from https://books.google.com/books?id=KXufePCNA5gC&pgis=1

Lynch, M., Cole, S. A., McNally, R., & Jordan, K. (2010). Truth Machine: The Contentious History of DNA Fingerprinting. University of Chicago Press. Retrieved from https://books.google.com/books?id=EVmr3kyB7ewC&pgis=1

Lynch, V. (2012). Solving crime with DNA What is the real story around DNA and crime scenes ?, 8(2), 3–9.

Martin, M., Carrington, M., & Mann, D. (1992). A method for using serum or plasma as a source of DNA for HLA typing. Human Immunology, 33(2), 108–113.

McCartney, C. (2013). Forensic Identification and Criminal Justice. Routledge. Retrieved from https://books.google.com/books?id=dWQQBAAAQBAJ&pgis=1

McCaskie, T. C. (2003). State and society in pre-colonial Asante (Vol. 79). Cambridge University Press.

McEwen, J. E. (1995). Forensic DNA data banking by state crime laboratories. American Journal of Human Genetics, 56(6), 1487.

Meyerowitz, E. L. R. (1952). Akan traditions of origin. Faber & Faber.

Moretti, T. R., Ph, D., Baumstark, A. L., Defenbaugh, D. A., Tr, R. M., Al, B., & Da, D.

(2001). Validation of Short Tandem Repeats (STRs) for Forensic Usage : Performance Testing of Fluorescent Multiplex STR Systems and Analysis of Authentic and Simulated Forensic Samples *, 647–660.

Moseley, R. (2011). Dna Fingerprinting. Worcester Polytechnic Institute.

Mozayan Mozayan, N. C. (Ed.). (2010). The Forensic Laboratory Handbook Procedures and Practice. Springer Science & Business Media. Retrieved from https://books.google.com/books?id=Kqhj3ugmYhQC&pgis=1

Mpoloka, S. W., Kgotlele, T., & Wally, A. (2008). Determination of allele frequencies in nine short tandem repeat loci of five human sub-populations in Botswana. African Journal of Biotechnology, 7(8).

Murangira, T., & Jyoti, B. (2012). DNA Technology: The Technology of Justice-Current and Future Need. Research Journal of Recent Sciences ISSN, 2277, 2502.

National Institutes of Health,(2015) US National Library of Medicine, D. of H. & H. S.. What is DNA? Retrieved from http://ghr.nlm.nih.gov/handbook/basics/dna

National Research Council, F. S. (1996). The evaluation of forensic DNA evidence. National Academies Press (US).

Noyd, R., Krueger, J., & Hill, K. (2013). Biology: Organisms and Adaptations. Cengage Learning. Retrieved from https://books.google.com/books?id=eUkWAAAAQBAJ&pgis=1

Nyanteng Victor K., Peter Takyi Peprah, Lawford Boateng Acheamfuor&Emmanuel Nii Okang Tawiah. (2013). Report 2010 Population and Housing Report. Regional Analytical Report. Ashanti Region, 3.

Osman, A. E., Alsafar, H., Tay, G. K., Theyab, J. B. J. M., Mubasher, M., Sheikh, N. E., & Alharthi, H. (2015). Autosomal Short Tandem Repeat (STR) Variation Based on 15 Loci in a Population from the Central Region (Riyadh Province) of Saudi Arabia, 6(1), 6–10. http://doi.org/10.4172/2157-7145.1000267

Panayiotopoulos, C. P. (2010). A Clinical Guide to Epileptic Syndromes and their Treatment. Springer Science & Business Media. Retrieved from https://books.google.com/books?id=yJQQzPTcbYIC&pgis=1

Panneerchelvam, S., & Norazmi, M. N. (2003). Forensic DNA profiling and database. The Malaysian Journal of Medical Sciences : MJMS, 10(2), 20–6. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3561883&tool=pmcentrez&rendertype=abstract

Pláteník, J. (2009). Electrophoresis in biochemistry, pages 2–7.

Poetsch, M., El-Mostaqim, D., Tschentscher, F., Browne, E. N. L., Timmann, C., Horstmann, R. D., & Von Wurmb-Schwark, N. (2009). Allele frequencies of 11 X-chromosomal loci in a population sample from Ghana. International Journal of

Legal Medicine, 123(1), 81–83.

Pompanon, F., Bonin, A., Belleman, E., & Taberlet, p. (2005). Genotyping errors: causes, consequences and solutions. Nature reviews. Genetics, 6(11), 847.

Primorac, D., Schanfield, M. S., & Primorac, D. (2000). Ap pli ca tion of Fo ren sic DNA Testing in the Le gal Sys tem, 41(1), 32–46.

Pyrek, K. (2010). Forensic Science Under Siege: The Challenges of Forensic Laboratories and the Medico-Legal Investigation System. Academic Press. Retrieved from https://books.google.com/books?id=neqi1xH078EC&pgis=1

Rafferty, J. P. (2011). New Thinking about Evolution. The Rosen Publishing Group. Retrieved from https://books.google.com/books?id=iP7HCYLRELsC&pgis=1

Reference Manual on Scientific Evidence. (2011). Washington, D.C.: National Academies Press. http://doi.org/10.17226/13163

Rettner, R. (2013). DNA: Definition, Structure & Discovery | What Is DNA? Retrieved November 9, 2015, from http://www.livescience.com/37247-dna.html

Robert D. Stevens, Nicholas Hart, M. S. H. (2014). Textbook of Post-ICU Medicine: The Legacy of Critical Care. Oxford University Press. Retrieved from https://books.google.com/books?id=aL6LAwAAQBAJ&pgis=1

Roewer, L. (2013). DNA fingerprinting in forensics: past, present, future. Investigative Genetics, 4(1), 22. http://doi.org/10.1186/2041-2223-4-22

Rudin, N., & Inman, K. (2001). An introduction to forensic DNA analysis. CRC press.

Sergio, P. D. (2013). DNA Fingerprinting: State of the Science. Springer. Retrieved from https://books.google.com/books?id=44EiBAAAQBAJ&pgis=1

Shewale, J. G., & Liu, R. H. (2013). Forensic DNA Analysis: Current Practices and Emerging Technologies. CRC Press.

Shrivastava, P., Jain, T., & Trivedi, V. Ben. (2015). Genetic polymorphism study at 15 autosomal locus in central Indian population. SpringerPlus, 4(1), 566. http://doi.org/10.1186/s40064-015-1364-1

Sinden, R. R. (2012). DNA structure and function. Elsevier.

Singleton, P. (2012). Dictionary of DNA and Genome Technology. John Wiley & Sons. Retrieved from https://books.google.com/books?id=cu_0Bek5j34C&pgis=1

Strachan, T., Goodship, J., & Chinnery, P. (2014). Genetics and Genomics in Medicine. Taylor & Francis. Retrieved from https://books.google.com/books?id=oRfhAwAAQBAJ&pgis=1

Sun, J. X., Helgason, A., Masson, G., Ebenesersdóttir, S. S., Li, H., Mallick, S., … others. (2012). A direct characterization of human mutation based on

microsatellites. Nature Genetics, 44(10), 1161–1165.

Talwar, P. (2014). Manual of Cytogenetics in Reproductive Biology. JP Medical Ltd. Retrieved from https://books.google.com/books?hl=en&lr=&id=-c8iAwAAQBAJ&pgis=1

Tamarin, R. H. (2015, March 20). Principles of Genetics. McGraw-Hill. Retrieved from http://localhost:8080/xmlui/handle/123456789/478

Taupin, J. M. (2013). Introduction to Forensic DNA Evidence for Criminal Justice Professionals. CRC Press. Retrieved from https://books.google.com/books?id=2rfMBQAAQBAJ&pgis=1

Therman, E., & Susman, M. (2012). Human Chromosomes: Structure, Behavior, and Effects (third edit). Springer Science & Business Media. http://doi.org/10.1007/978-1-4684-0529-3

Thiele, K., Löffler, S., Löffler, J., Günthner, F., Nitschke, K., Edelmann, J., & Lessig, R. (2008). Population data of eight X-chromosomal STR markers in Ewe individuals from Ghana. Forensic Science International: Genetics Supplement Series, 1(1), 167–169. http://doi.org/10.1016/j.fsigss.2007.10.059

Ubelaker, D. H. (2012). Forensic Science: Current Issues, Future Directions. John Wiley & Sons. Retrieved from https://books.google.com/books?id=pOtjI8mIPSoC&pgis=1

White, H. W., Kusukawa, N., & Bioproducts, F. M. C. (1997). Agarose-Based System for Separation of Short Tandem Repeat Loci, 22(5), 976–980.

Yasin, S. R., Hamad, M. M., ElKarmi, A. Z., & Jaran, A. S. (2005). African Jordanian Population Genetic Database on Fifteen Short Tandem Repeat Genetic Loci. Croatian Medical Journal, 46(4), 587.

Yazdi, P. G., Pedersen, B. A., Taylor, J. F., Khattab, O. S., Chen, Y.-H., Chen, Y. & Wang, P. H. (2015). Nucleosome Organization in Human Embryonic Stem Cells. Plos One, 10(8), e0136314. http://doi.org/10.1371/journal.pone.0136314

Ziętkiewicz, E., Witt, M., Daca, P., Żebracka-Gala, J., Goniewicz, M., Jarząb, B., & Witt, M. (2012). Current genetic methodologies in the identification of disaster victims and in forensic analysis. Journal of Applied Genetics, 53(1), 41–60. http://doi.org/10.1007/s13353-011-0068-7

APPENDIX

1. QUESTIONNAIRE

FREQUENCY DISTRIBUTION OF SELECTED STR MARKERS AMONG THE ASHANTI POPULATION OF GHANA.

CODE: |__|__|__|

INTERVIEW INFORMATION

DATE OF INTERVIEW |__|__| Day |__|__| Month|__|__|__|__|

Year

RESPONDENT'S IDENTIFICATION
RESPONDENT'S TOWN OF RESIDENCE |__|__|__|__|__|__|__|__|
RESPONDENT PHONE NUMBER |__|__|__|__|__|__|__|__|__|__|

INTERVIEWER: INTRODUCTION AND CONSENT. May I begin the interview now?

QUEST. NO.	QUESTIONS AND FILTERS	CODING CATEGORIES	SKIP

Q1.	Consent	1. Yes 2. No	
Q2.	Sex	1. Male 2. Female	
Q3.	Please tell me your date of birth or your age in years.	|__|__| day |__|__| month |__|__|__|__| year |__|__| age (completed years) Don't know……..	
Q4.	Which region do you come from?		
Q5.	Which district do you come from?		
Q6.	To which ethnic group do you belong?		
Q7.	Which clan do you belong to?	1. Oyoko 5. Ekuona 2. Agona 6. Asakyiri 3. Aduana 7. Asenie 4. Bretuo 8. Asona 9. Others (specify) ……	

Q8.	What is your mother's name?		
Q9.	Which region does she come from?		
Q10.	Which district is she from?		
Q11.	To which ethnic group does she belong to?		
Q12.	To which clan does she belong to?	1. Oyoko 2. Agona 3. Aduana 4. Bretuo 5. Ekuona 6. Asakyiri 7. Asenie 8. Asona 9. Others (specify) …….	
Q13.	What are the names of your mother's parents?	a. Grandmother…… b. Grandfather……	
Q14.	Which region do they come from?	a. Grandmother……	

		b. Grandfather……	
Q15.	Which district do they come from?	a. Grandmother…… b. Grandfather……	
Q16.	To which ethnic groups do they belong to?	a. Grandmother……. b. Grandfather……	
Q17.	To which clan do they belong to?	a. Grandmother……. b. Grandfather……	
Q18.	What is your father's name?		
Q19.	Which region does he come from?		
Q20.	Which district is he from?		
Q21.	Which ethnic group does he belong to?		

Q22.	Which clan does he belong to?	1. Oyoko 5. Ekuona 2. Agona 6. Asakyiri 3. Aduana 7. Asenie 4. Bretuo 8. Asona 9. Others (specify)……	
Q23.	What are the names of your father's parents?	a. Grandmother……. b. Grandfather……	
Q24.	Which region do they come from?	a. Grandmother……. b. Grandfather……	
Q25.	Which district do they come from?	a. Grandmother……. b. Grandfather……	
Q26.	To which ethnic groups do they belong to?	a. Grandmother……. b. Grandfather……	
Q27.	Do you have any unique family name?	1) YES	

		2) NO	
Q28.	What is the family name?		

Thump prints

Left thump

Right thump

INTERVIEWER'S REMARKS		
I1.Was anyone else present during the interview? No one else was present .....1 Respondent's doctor/nurse ....2 Other clients/Patients .....3 Respondent's relatives.....4 Other Health professionals 5	**I2.In general, what was the respondent's attitude during the interview?** Friendly, interested..... 1 Was cooperative, but not particularly interested..... 2 Impatient, worried 3	**I3.Did the respondent understand the questions?** Understood well.....1 Did not understand very well.2 Understood poorly 3

	Hostile 4	

2. CONSENT FORM

Statement of person obtaining informed consent:

I have fully explained this research to _____ and have given sufficient information about the study, including that on procedures, risks and benefits, to enable the prospective participant make an informed decision to or not to participate.

DATE: _____ NAME: _____

Statement of person giving consent:

I have read the information on this study/research or have had it translated into a language I understand. I have also talked it over with the interviewer to my satisfaction.

I understand that my participation is voluntary (not compulsory).

I know enough about the purpose, methods, risks and benefits of the research study to decide that I want to take part in it.

I understand that I may freely stop being part of this study at any time without having to explain myself.

I have received a copy of this information leaflet and consent form to keep for myself.

NAME:_____

DATE: _____ SIGNATURE/THUMB PRINT: _____

Statement of person witnessing consent (Process for Non-Literate Participants):

I _____ (Name of Witness) certify that information given to

_____(Name of Participant), in the local language, is a true reflection of what l have read from the study Participant Information Leaflet, attached.

WITNESS' SIGNATURE (maintain if participant is non-literate): _____

MOTHER'S SIGNATURE (maintain if participant is under 18 years): _____

MOTHER'SNAME: _____

FATHER'S SIGNATURE (maintain if participant is under 18 years): _____

FATHER'S NAME: _____

3) FORMULAR FOR THE CALCULATION OF POLYMORPHISM INFORMATION CONTENT (PIC).

PIC =

$$1 - \sum_{i=1}^{n} p_i^2 - \left(\sum_{i=1}^{n} p_i^2 \right)^2 + \sum_{i=1}^{n} p_i^4$$

CALCULATION OF FORENSIC EFFIECIENCY PARAMETERS.

A). FGA marker. B). D18S51 marker.

FGA				D18S51			
Alleles	freq.	Freq.2	Freq.^{4}	Alleles	freq.	Freq.2	Freq.^{4}
25.1	0.01265823	0.000160231	2.56739E-08	8.1	0.015384615	0.000236686	5.60204E-08
26.1	0.0379747	0.001442077	2.07958E-06	10	0.061538462	0.003786982	1.43412E-05
27.2	0.0506329	0.002563692	6.57252E-06	14.1	0.061538462	0.003786982	1.43412E-05
28.3	0.0253165	0.000640923	4.10782E-07	16.2	0.123076923	0.015147929	0.00022946
30.1	0.0126582	0.000160231	2.56739E-08	19	0.2	0.04	0.0016
31.2	0.0253165	0.000640923	4.10782E-07	21.2	0.169230769	0.028639053	0.000820195
33	0.1139241	0.012978689	0.000168446	24.1	0.061538462	0.003786982	1.43412E-05
33.3	0.0126582	0.000160231	2.56739E-08	27.1	0.2	0.04	0.0016
34.2	0.1139241	0.012978689	0.000168446	30.2	0.107692308	0.011597633	0.000134505
36.1	0.2405063	0.057843294	0.003345847				
38	0.1518987	0.023073225	0.000532374	Total	1.000000001	0.146982249	0.00442724
39.3	0.1139241	0.012978689	0.000168446				
41.2	0.0506329	0.002563692	6.57252E-06				
43.2	0.0379747	0.001442077	2.07958E-06				
Total	1	0.129626663	0.004401762				
(SUM of Freq.2)2 =		0.016803072		(SUM of Freq.2)2 =		0.021603781	
Homozygossity of FGA=		0.129626663		Homozygossity of D18S51 =		0.146982249	
Expected Heterozygosity=		0.870373337		Expected Heterozygosity=		0.853017751	
Polymorphism Information Content (PIC)of FGA =		0.857972028		Polymorphism Information Content (PIC) of D18S51 =		0.83584121	