Application of autosomal INDELs as a forensic tool in Qatar

By

Majid Bashir

A thesis submitted in partial fulfillment for the requirements for the degree of PhD at the University of Central Lancashire

June 2016

STUDENT DECLARATION FORM

Concurrent registration for two or more academic awards

I declare that while registered as a candidate for the research degree, I have not been a registered candidate or enrolled student for another award of the University or other academic or professional institution

---

Material submitted for another award

I declare that no material contained in the thesis has been used in any other submission for an academic award and is solely my own work

Signature of Candidate 

Type of Award **PhD**

School **Forensic and Investigative Sciences**

Abstract

Short tandem Repeats (STRs) are the most commonly genetic markers used in forensic human identification. However, in some cases they are not able to yield complete profiles because of DNA degradation and/or inhibition. The STR profiling of the degraded/inhibited DNA samples can result in allelic drop-outs and even no profile at all. Alternatively, single nucleotide polymorphisms (SNPs) can be used to address the issues of DNA degradation and inhibition due to their smaller amplicons. But their use in regular forensic case work is limited due to additional steps (sequencing based) and time consuming process. To bridge the gap between STRs and SNPs by combining their characteristics, another type of genetic marker, insertion deletion polymorphisms (INDELs), offer an effective way to analyze challenged DNA sample (degraded and inhibited ones). INDELs have short amplicons, low mutation rates, no stutter peaks and are analyzed using the same equipment and protocols as STR polymorphisms.

In this study, the forensic efficiency of 30 autosomal INDELs contained within the Qiagen™ Investigator™ DIPplex kit were tested by using 500 samples from individuals belonging to five different nationalities (Qatari, Pakistani, Sudanese, Tunisian and Yemeni) based in Qatar. Population indices and forensic parameters were calculated. The results showed no significant deviation from Hardy-Weinberg Equilibrium and no evidence of linkage disequilibrium for all of 30 markers was found after applying Bonferroni's correction ($P < 0.00166$). The Combined Power of Discrimination (CPD) for the 30 INDEL loci was 0.9999999 for all of five populations which shows that these 30 markers are very efficient and suitable for forensic casework. The Combined Probability of Match (CPM) was calculated in the order of 10^{-13} which is a satisfactory value for forensic purposes. In addition to assess forensic efficiency of 30 autosomal INDELs, an effort was also made to derive ancestry information by using different statistical systems (Arlequin, Snipper and STRUCTURE). The results indicated that 30 INDELs contained in Qiagen™ Investigator® DIPplex kit, although useful for forensic identification, poorly differentiate five population groups of Qatar. The reason of failure in differentiating the populations, lies in the selection of INDEL markers, which were chosen for identification purpose (i.e. they have similar allele frequencies in different populations) rather than deriving the ancestry information (i.e.

ancestry informative markers are chosen on the basis that their allele frequencies need to be different for different populations).

In order to recover genetic information lost during DNA degradation, the concept of reduced sized PCR products (mini amplicons) has been developed. MiniFiler® kit (Applied Biosystems™) containing 8 mini-STRs (developed by re-designing the primers of 8 STRs contained in Identifiler Plus® kit) provide a possible solution of DNA degradation. By using similar mini-plex approach, a multiplex PCR assay for INDELs (named as mini-INDELs) has been developed in this research. A total of 14 autosomal INDEL markers were selected and short amplicons were designed for them keeping in view that they could perform efficiently on degraded samples. The multiplex of 14 INDELs was designed and optimized successfully in a single tube reaction. All the markers were amplified adequately with good peak balance and expected amplicon sizes. The sensitivity of mini-INDELs was found upto 0.03125 ng of genomic DNA with complete and balanced profile. The concordance between mini-INDELs kit and Qiagen™ Investigator™ DIPplex kit (for the common loci) was observed in 99.7% INDEL alleles. The efficiency of mini-INDELs PCR assay was also evaluated on a set of artificially prepared degraded and inhibited DNA samples.

It is concluded that INDELs markers in general and mini-INDELs in specific, can be used as a useful tool in forensic case work and can also be employed in conjunction with STR typing as a complementary tool especially in those cases where low level of DNA and DNA degradation are suspected.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ADD: Accumulated Degree Days

AFLP: Amplified Fragment Length Polymorphism

AIMs: Ancestry Informative Markers

bp: Base Pair

CPD: Combined Power of Discrimination

CPE: Combined Probability of Exclusion

CPI: Combined Probability of Identity

CPM: Combined Power of Match

°C: Degrees Celsius

DIP: Deletion Insertion Polymorphism

DNA: Deoxyribonucleic Acid

g: Gram

He: Expected Heterozygosity

Ho: Observed Heterozygosity

HWE: Hardy-Weinberg Equilibrium

INDEL: Insertion Deletion polymorphism

µg: Microgram

µl: Microliter

mM: Millimolar

ml: Milliliter

mtDNA: Mitochondrial DNA

ng: Nanogram

NGS: Next Generation Sequencing

P-Value: Probability value

PCA: Principal Component Analysis

PCI: Phenol Chloroform Isoamyl-alcohol

PCR: Polymerase Chain Reaction

PD: Power of Discrimination

PE: Power of Exclusion

PIC: Polymorphism Information Content

PM: Probability of Match

RFU: Relative Fluorescence Units

RFLP: Restriction Fragment Length Polymorphism

STR: Short Tandem Repeat

SNP: Single Nucleotide Polymorphism

TPI: Typical Paternity Index

List of Tables

List of Figures

ACKNOWLEDGEMENT

All thanks to Allah, Who protect and bless me throughout my life.

First of all, I would like to thank my supervisor, Dr. William Goodwin who has provided me with kindness, guidance, co-operation and advice throughout the course of my PhD project. I really appreciate his efforts to convert me from a forensic practitioner to a forensic researcher. Also, I would like to thank Dr. Sibte Hadi, Dr. Arati Iyengar, Dr. Judith Smith and Professor Jaipal Singh for their advice and support.

I am highly grateful to my late father, Muhammad Bashir, whose vision inspires me to achieve the highest degree. My mother remains a constant source of blessing for me and her prayer enables me to achieve a lot in this world.

I appreciate the support and co-operation of my wife, Gulnaz. It would be difficult for me to go through if she would not take special care of me and my kids during the period of studies. I am also thankful to my kids, Hussain, Huda, Heba and Haya who motivate me to give them a highly qualified father.

I am grateful to my friends and colleagues at Qatar lab and UCLAN for their help during the period of my PhD.

Finally, very big thanks go to my elder brother like boss, Brigadier Ibrahim Abdulla Al-Binali whose trust, encouragement, motivation, support and kindness helped me to achieve a lot of my goals of my life along with this degree.

CHAPTER 1
INTRODUCTION

1.1 Historical prospective of forensic analysis

The use of genetic markers for forensic purposes, started with the implementation of ABO blood grouping (Landsteiner 1900). These classical markers provided a starting point in serology based human identification, but due to their limited power of discrimination (1 in 10) and instability in normal environmental conditions, their use was limited (Budimlija *et al.* 2003). Today, 30 major blood group systems have been recognized by the International Society of Blood Transfusion. A number of genes and alleles which contribute blood group antigens have been characterized. However, in the recent times, the use of serological is limited to medical applications.

The use of protein markers has also been attempted in forensic human identification. The sequences of amino acids in some proteins differ among individuals. Isoenzymes, present on human red blood cells as well as blood serum, are enzymes with multiple forms which can catalyze the same chemical reaction in spite of different sequences of amino acids. Due to this property, they can act as genetic markers which can differentiate between people (Budowle *et al.* 1985). The techniques used to separate protein markers into distinct alleles, involved starch gel, agarose gel and polyacrylamide gel electrophoresis. Although providing an improved power of discrimination (1 in 100) as compared to ABO blood grouping, protein markers were still limited.

The use of DNA in human identification was first reported in 1985 by an English scientist named Alec Jeffreys. He discovered the repetitive regions of DNA and found that these regions of DNA differ from individual to individual. These DNA repeat regions were called VNTRs which stands for variable number of tandem repeats. Dr. Jeffreys examined VNTRs by using a technique called restriction fragment length polymorphism (RFLP). This RFLP method involved the use of a restriction enzyme to cut the regions of DNA surrounding the VNTRs (Jeffreys *et al.* 1985;

Nakamura *et al.* 1987). These DNA fragments of different lengths were separated on an agarose gel. Following DNA fragments separation on the agarose gel, a Southern blot was used to transfer the DNA fragments from gel to a nylon membrane where alleles were detected using radioactive probes (Southern 1975). The use of multi locus VNTRs was limited due to time required for their analysis (several days to weeks) and requirement of large amount of DNA (>50 ng), which is normally not available from crime scene samples (Patzelt 2004).

Single locus VNTR probes were adopted in the next phase. Each of single locus probes had a large number of alleles and had heterozygosities of around 90%. Several loci of single locus probes could be analyzed to generate a genetic profile, which helped in inclusion and exclusion of individuals during forensic DNA analysis (Jarman *et al.* 1986; Wong *et al.* 1986; Wong *et al.* 1987*).* The limitations of single locus probes were their laborious and time consuming process, inability to type degraded samples and difficulty in interpretation of mixtures (Jobling *et al.* 2004).

1.2 Modern Forensic DNA markers

The advent of polymerase chain reaction (PCR) revolutionized the forensic DNA testing. PCR can amplify a target sequence of DNA several million times and it is therefore possible to generate a profile by using minute quantities of DNA (Mullis *et al.* 1986; Kline *et al.* 2005). PCR reactions make the back bone of recent forensic DNA analysis.

1.2.1 Short Tandem Repeats (STRs)

STRs also called microsatellites, are DNA repeat regions of 2-6 base pair in length and occur throughout human genome on average every 10,000 nucleotides (Collins *et al.* 2003). STRs required only small amounts of DNA (< 1 ng) to generate a DNA profile and due to the ability to co-amplify several polymorphic markers in the same PCR, they are both highly sensitive and discriminating when used for human identification (Saiki *et al.* 1988; Schmerer *et al.* 1999; Butler 2003). A number of commercial STR kits have been developed and are widely used by forensic scientists. The most recent STR kits used in forensic labs are GlobalFiler® (Applied Biosystems™), PowerPlex® Fusion (Promega™) and Investigator® 24plex QS (Qiagen™).

Although STRs are still the most widely used genetic markers, they have some shortcomings, which include their relatively large amplicon sizes (150- 450 bp), which can result in poor or failed profiling for degraded and inhibited samples (Li *et al*. 2011; Manta *et al*. 2012), the presence of stutters which can make analysis complex, especially in the case of mixture interpretation (LaRue *et al*. 2012). In addition, their relatively high mutation rates can complicate kinship analysis (Weber *et al.* 2002).

1.2.2 Y-chromosomal short tandem repeats (Y-STRs)

STRs markers on the Y-chromosome are another type of genetic markers (lineage markers) used in forensic DNA analysis and is inherited from father (Butler 2006).

The latest commercial kits available for Y-STRs typing include Y-Filer® Plus (Applied Biosystems™) and Powerplex® Y23 (Promega™). Y-STRs are especially helpful in sexual assault cases where male DNA can be differentiated even in the presence of many fold amounts of female DNA (Kayser 2007). Other applications include human migration and evolutionary studies, verifying amelogenin Y deletion, paternity cases especially where male children can be linked to a putative paternal relative, which can be valuable when the mother's sample is not available or in fatherless paternity cases where comparisons can be made with male relatives (Thangaraj *et al.* 2002; Kayser *et al.* 2007; Marjanovic *et al.* 2009). Another class of STR markers on the Y-chromosome is rapidly mutating Y-STRs, which have the ability to differentiate between close or distant male relatives (Ballantyne *et al.* 2012). Although Y-STRs are useful in

certain cases, its widespread use is limited due to its haplotype nature and limited discriminatory power (Palo *et al.* 2007).

1.2.3 X-chromosomal short tandem repeats (X-STRs)

The X-chromosome is one of the two sex chromosomes in humans. It is the only chromosome which has one of the pair inactivated in females (Gunter 2005). The X-chromosome is 153Mb in base length and constitutes 5% of female genetic material and 2.5% of males who have single copy (Ross *et al.* 2005). The commercial kit available for typing X-STRs although with limited use is Investigator® Argus X-12 kit (Qiagen™).

The use of X-chromosome STR tying helps in kinship scenarios which involve at least one female or in motherless cases, half-sister testing using father as a reference, grandparent-grandchild comparisons and paternity testing in incest cases (Barbaro *et al*. 2006d; Toni *et al*. 2003; Toni *et al.* 2006).

1.2.4 Mitochondrial DNA

Human mitochondrial DNA (mtDNA) is also classified as lineage marker like Y-chromosome polymorphisms and is maternally inherited. It has a size of approximately 16,569 bp and most of sequence variations are found in two hypervariable (HV) regions called HV-I and HV-II (Holland and Parson 1999). Due to the fact that approximately 1000 copies of mtDNA are present in each cell and the molecules show some resistance to degradation due to their circular nature (Salas *et al*. 2007), mtDNA became a powerful tool to type highly degraded forensic samples (Isenberg 2004, Nelson and Melton 2007, Sudoyo *et al.* 2008). However, due to multiple steps required to generate profiles and the complexities in interpretation, mtDNA analysis is limited in regular forensic DNA analysis (Holland *et al*. 1999, Asari *et al.* 2007).

1.2.5 Single Nucleotide Polymorphisms (SNPs)

A single base sequence variation at a particular point in the genome is called single nucleotide polymorphism (SNP). SNPs are widely spread in human genome once every 1000 bases (Venter 2001). Several groups and institutions such as SNP consortium (Thorisson and Stein 2003; Butler 2005) have worked to map and characterize huge number of SNPs through different projects like HapMap project (Thorisson *et al.* 2005; Frazer *et al.* 2007; Altshuler *et al.* 2010; Jobling *et*

al. 2013) and have developed SNPs databases including HGDP-CEPH Diversity panel and dbSNPs (Cann *et al.* 2002; Rosenberg *et al.* 2002; Cavalli-Sforza 2005; Li *et al.* 2008; Fondevila *et al.* 2013). SNPs have lower mutation rate than those of STRs (Nachman and Crowell 2000) which make them ideal to study human evolutionary pattern and to derive ancestry information (Brion *et al*. 2005). Due to the presence of smaller amplicons, SNPs had been shown to be effective when typing degraded DNA (Dixon *et al*. 2006; Alaeddini *et al*. 2010).

The major shortcoming of SNPs is that they are less discriminatory than STRs and about 50 SNPs are required to have same discrimination power as 12 STRs (Gill *et al.* 2001). In addition to less discrimination power, the methodology used for SNP analysis has multiple steps and interpretation can be challenging, especially when mixtures are present (Philips *et al*. 2007).

1.2.6 Insertion Deletion Polymorphisms (INDELs)

In recent years, another type of genetic markers called INDELs (Insertions-Deletions) has been getting attention from forensic scientists. This is due to the presence of the combined advantages of STRs and SNPs:, wide distribution in human genome, their lower mutation rates and the use of smaller amplicons and routine genotyping techniques (Li *et al.* 2011; Manta *et al*. 2012; Borsting *et al.* 2013; Cereda *et al*. 2014). These characteristics make them ideal candidates for forensic applications, especially for the profiling of degraded and inhibited samples (Romanini *et al*. 2012).

1.2.6.1 INDEL formation

Several mechanisms have been hypothesized for the formation of INDELs which include the slippage of DNA strands during replication or repair as short repetitive sequences (Levinson *et al.* 1987) and deletion in loop regions of DNA secondary structure (Hoot and Douglas 1998; Vom and Hachitel 1988). The expansion, contraction and diversification of these INDEL repeat regions can happen among different lineages in multiple ways and these INDEL sequences can also be independent of flanking regions (Golenberg *et al.* 1993; Benson and Dong 1999) .Other genetic elements such as *Alu*, L1 and SVA can also be responsible for INDEL variations in humans. *Alu* element can produce insertions of about 300 bp (Comas *et al*. 2000), while L1 and SVA retro transposons can cause insertions in the range of 10 bp to 3 kb in the human genome

(Bennett *et al.* 2004; Mills *et al*. 2007). Collectively, these elements cause millions of insertions in human populations (Beck *et al*. 2010; Huang *et al*. 2010). Figure 1.1 illustrates different mechanisms of INDELs formation.

1.2.6.2 Identification of INDELS

The first large-scale efforts to identify INDELs in the human genome were focused on human chromosome 22 and the re-sequencing data from 31 humans for this particular chromosome identified 13% of the genetic variations as INDEL polymorphisms (Mullikin *et al*. 2000; Dawson *et al.* 2001). Later on, several other publications revealed the identification of more INDELs:

- The identification of 2393 INDELs (1 bp to 543 bp) was carried out in diverse populations through the re-sequencing of 330 genes (Bhangale *et al*. 2005).
- In a follow-up study, PolyPhred ver. 6.0 was used to identify 1126 additional INDELs in the ENCODE regions of the human genome (Bhangale *et al*. 2006).
- The mapping of 415,000 unique INDELs, with an average density of one INDEL per 7.2 kb, was done using a new computational strategy (Mills *et al.* 2006).
- A similar re-sequencing technique called PolyScan was developed and used for heterozygous INDELs detection (Chen *et al*. 2007).
- 796,273 small INDELs were reported in a project that was designed to detect structural variation in eight diverse humans (Kidd *et al.* 2008).
- Recent follow-up studies have reported around 2 million small insertions/deletions ranging from 1 bp to 10,000 bp in length in the genomes of 79 diverse humans on 22 autosomes and X and Y chromosomes (Mills *et al.* 2011). About 41% of them are apparently random DNA sequences and most of them are under 100 bp (Neuvonen *et al*. 2011; Friis *et al*. 2012; Montgomery *et al*. 2013).

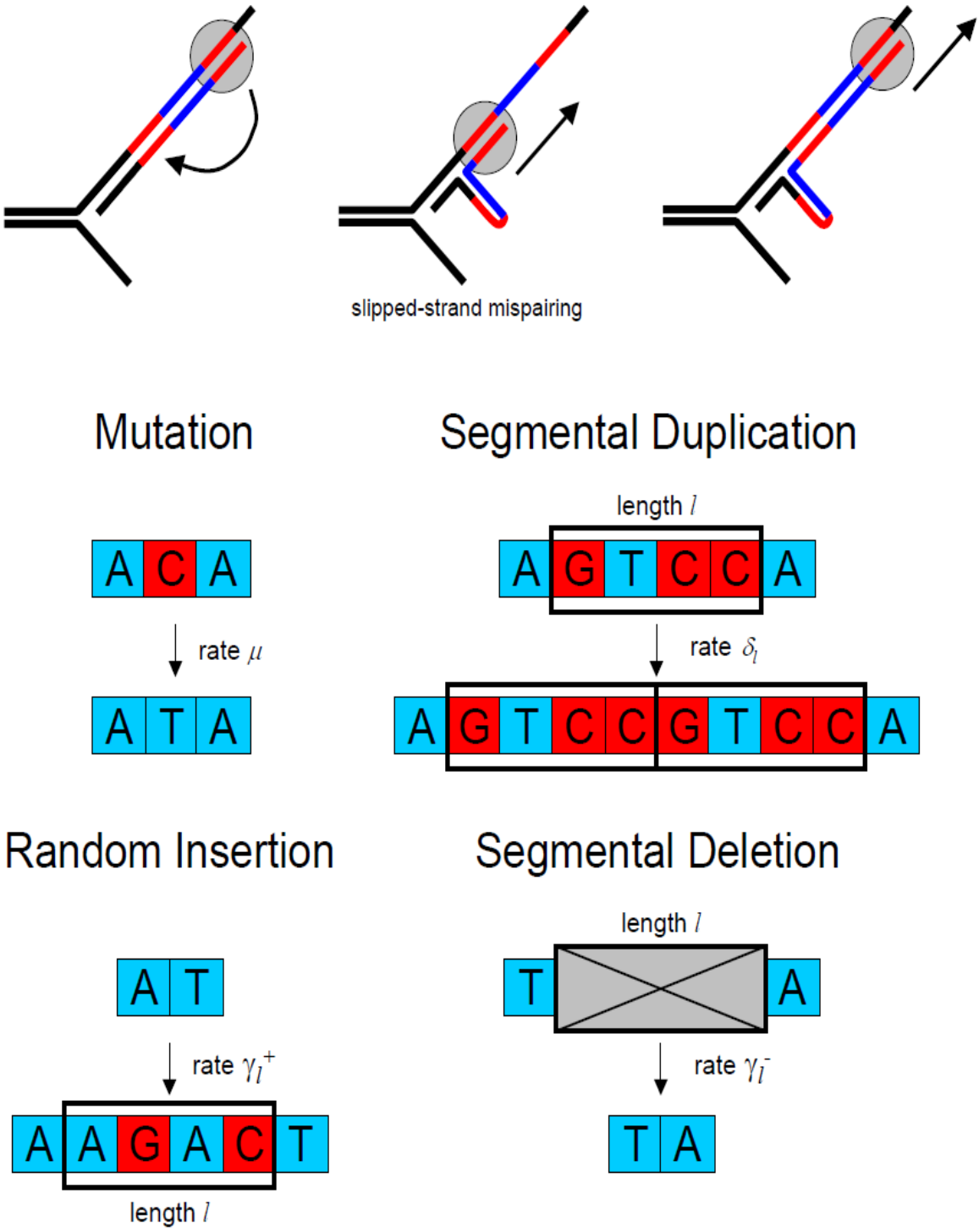

Figure.1.1 Diagram showing different mechanisms of INDELs formation (adopted from Messer P.W 2007).

1.2.6.3 Applications of INDELS

1.2.6.3.1 Forensic Application

The forensic application of INDELs has received attention in the forensic community. The major use in this regard remains the recovery of genetic information from challenged samples, i.e. degraded and inhibited (Pereira *et al.* 2012). They are also useful for the interpretation of mixtures due to the absence of microvariants products (Carvalho *et al*. 2011).

1.2.6.3.2 Kinship analysis

Kinship analysis is another field of interest where INDELs have been applied. In majority of cases, STRs can clearly resolve paternity cases either as inclusion or exclusion due to high levels of polymorphism (Amorim *et al*. 2005). But there are some scenarios when statistical results from STRs are ambiguous. For example, the paternity cases in which the alleged father might be a close relative of the real father (Karlsson *et al.* 2007). In these cases, bi-allelic markers such as SNPs and INDELs are the alternative choices due to their low mutation rates as compared to STRs (Phillips *et al*. 2007; Pereira *et al*. 2009; Borsting *et al.* 2011).

1.2.6.3.3 INDELS as Ancestry Information Markers (AIMs)

The main application of Ancestry Information Markers (AIMs) in population genetics is to determine ancestry proportions and structures in admixed populations. In forensic genetics, AIMs can be used as a useful tool to derive the possible ancestry of the forensic evidence (Fridakis *et al.* 2003; Phillips *et al*. 2007; Kayser and De Kniff 2011). To predict the biogeographic ancestry of unknown evidential samples, STRs are limited due to their high mutation rates and similar alleleic distributions in different populations, which can result in wrong assignment of population (Phillips *et al.* 2007). SNPs have a number of characteristics which make them ideal to get information for ancestry. They have low rate of mutation, high distribution in the genome and varying allele frequency patterns across populations (for some SNPs). However, due to the requirement of multiple steps for their typing and complex interpretation of their data, their use become limited in the routine forensic laboratories (Fridakis *et al.* 2003; Phillips *et al.* 2007; Fondevila *et al.* 2008).

Like SNPs, INDELs have also low mutation rate (Nachman and Crowell 2000) and they have high distribution of allele frequency among distant populations at some loci (Weber *et al*. 2002). These factors make them ideal candidates as ancestry informative markers. They can also be used to identify sub-structure in mixed populations (Santos *et al*. 2010). Different multiplexes for autosomal INDELs were prepared to evaluate distribution and ancestry information of diverse populations (Santos *et al*. 2010; Pereira *et al.* 2012; Francez *et al.* 2012; Cardena *et al.* 2013; Manta *et al*. 2013; Zaumsegel *et al*. 2013; LaRue *et al*. 2014; Romanini *et al.* 2014).

1.2.6.4 Development of INDELs multiplexes

Several INDEL multiplexes have been developed including in-house (Phillips *et al*. 2007; Pereira *et al*. 2009; Fondevila *et al.* 2011; Kis *et al*. 2012) and one commercial kit for autosomal markers (Qiagen™ Investigator® DIPplex). This kit was commercially launched in 2009 and it allows simultaneous amplification of 30 biallelic Deletion/Insertion Polymorphisms (DIPs) which spread over 19 autosomes and are separated by at least 10 Mb from commercially available STRs loci plus a specific fragment of the amelogenin gene as a sex marker, with a maximum amplicon length of 160 bp (Qiagen™ Investigator® DIPplex handbook). Since the manufacture of the Investigator® DIPplex kit, a number of world populations (Argentinian, Bangladeshi, Brazilian, Chinese, German, Somalian, South Korean, Japanese) have been profiled (Larue *et al.* 2012; Manta *et al.* 2012; Martin *et al*. 2013; Akhteruzzaman *et al.* 2013; Seong *et al.* 2013; Pinto *et al*. 2013; Zaumsegel *et al.* 2013; Nunotani *et al.* 2015).

1.2.6.5 Statistical value of INDELS

The applicability of 30 autosomal INDELs (using Investigator® DIPplex kit) has been proved for forensic use and genetic diversity in number of populations (Table 1.1). A high level of power of discrimination has been observed in all of typed populations (> 0.9999). The value of combined match probability has been calculated in the range of 10^{-11} to 10^{-14} (Pinto *et al.* 2013; Zaumsegel *et al.* 2013). The mean typical paternity index (TPI) ratios are in the range of 0.800-1.00 (Seong *et al.* 2013; Nunotani *et al*. 2015). For most of the populations, no deviation from Hardy-Weinberg equilibrium was observed for any of INDEL markers (Larue *et al*. 2012; Manta *et al*. 2012; Martin *et al*. 2012; Pinto *et al.* 2013; Zaumsegel *et al.* 2013). Table 1.1 shows the

comparison of the values of combined power of exclusion and identity for different kits compared to Investigator® DIPplex kit.

Overall, it can be said that 30 INDELs can be used for forensic identification and kinship analysis.

Table 1.1 Table showing a comparison of CP values for different forensic human identification kits (Adapted from Qiagen™ investigator® DIPplex kit Handbook).

Kits	Loci	CPE/Trio*	CPI/CPM^	Population
Investigator DIPplex	30 DIPs	0.9980	2.83×10^{-13}	**German Caucasian**
AmpFlSTR Minifiler	8 STRs	0.999976	8.21×10^{-11}	**US Caucasian**
AmpFlSTR SEfiler plus	11 STRs	0.999998	7.46×10^{-14}	**US Caucasian**
Powerplex 16	15 STRs	0.9999994	5.46×10^{-18}	**US Caucasian**
Sanchez *et al*.2006	52 SNPs	0.9998	5.00×10^{-21}	**European Caucasian**

* Combined probability of exclusion

^ Combined probability of identity

^ Combined probability of match

1.3 Bio-geographical ancestry analysis

Bio-geographical ancestry analysis is based on human genomic variation which can help in deriving the geographical information of a particular population (Rosenberg *et al*. 2002). Bio-geographical analysis can help in estimating the ancestry of the donor of crime scene stains, identifying missing persons or disaster victims and confirming donor's self-declared ancestry in certain cases (Rohlfs *et al.* 2012).

The genetic markers used to derive ancestry information include lineage markers (Y and mtDNA) and autosomal markers (STRs, SNPs and INDELs).

Y and mtDNA variations being haplotypes are preserved in lineages and can correlate strongly with geographical regions (Brion *et al*. 2005; Wetton *et al.* 2005). However, these linage markers can misrepresent an individual's ethnic origin when distant male/female lineages are inherited that have a common ancestry. As co-ancestry in an individual indicates population admixture which can wrongly predict an individual's ancestry (King *et al*. 2007). In addition, with lineage markers, there is a need for much larger databases (such as YHRD/EMPOP) to measure haplotype variation covering all the regions of the world uniformly (Willuweit and Roewer 2007; Parson *et al*. 2007).

SNPs are better predictors of ancestry than STRs due to their low mutation rates i.e. once in 10^8 generations while STRs have approximately one in thousands (Barnholtz-Sloan *et al*. 2005). A number of SNPs have been identified as ancestry informative markers that have alleles with large frequency differences among different population groups (Shriver and Kittles 2004). Use of tri-allelic SNPs in different multiplex also helps in deriving ancestry information effectively due to their high population differentiation (Westen *et al*. 2009).

The SNP*for*ID developed by several European forensic DNA communities, selected several sets of forensic SNPs and developed SNP multiplex assays. They published a 52-plex SNP and SNaPshot assay (Sanchez *et al*. 2006). They and others have also developed a 34-plex and 47-plex SNP assays using ancestry informative SNPs (Phillips *et al.* 2007; Kersbergen *et al*. 2009;

Fondevila *et al.* 2013). Population data were gathered to calculate SNP allele frequency for these assays.

A company DNAPrint (Sarasota, Florida) provided a DNA test for determining a person's ethnic origin with a panel of SNPs (Frudakis *et al.* 2008). The company targeted pigmentation and xenobiotic metabolism genes while selecting AIM-SNPs (Frudakis *et al*. 2003).

Next generation sequencing (NGS) will not only help in increasing SNP multiplexing ability but will also enhance ancestry panel's informativeness. By using NGS platform (Ion torrent, PGM sequencing by Applied Biosystems™ or Miseq sequencing by Illumina™), a large amount of genotype data can be generated and population databases can be expanded.

Two approaches have been tried to derive ancestry information using autosomal STRs i.e. either by utilizing existing STR markers or employing specialist STRs with better population differentiation. The efforts have been made out to assess the ability of Identifiler Plus® (Applied Biosystems™) based 15 STRs for ancestry-informativeness (Londin *et al.* 2010; Phillips *et al.* 2011) but failed to differentiate the global sample set of 7 populations. But when 36 novel STRs including 33 dinucleotide-repeat STRs, were tried, better ancestry information was obtained (Pereira *et al.* 2012). To derive ancestry information, di-nucleotide repeat STR loci were found better than those of tri- or tetra-nucleotide repeat loci because they were much more differentiated across populations. Although di-nucleotide STRs are not ideal for forensic identification purpose due to their high percentage of stutter (usually 30 %) and (Walsh *et al.* 1996), they are better in ancestry inference than tetra-nucleotide repeat STRs. Several online browsers are available which contain databases of different world populations and can give an estimate of ancestry; pop.STR (http://spsmart.cesga.es/popstr.php), Snipper (http:// mathgene.usc.es/snipper) PopAffiliator2 (http://cracs.fc.up.pt/nf/popaffiliator2), where STR genotypes can be input and assignment probabilities for the population are returned.

INDELs like SNPs are bi-allelic genetic markers and can provide ancestry information. The Marshfield linkage marker sets (http://www.marshfieldclinic.org/mgs/pages/default.aspx?page5didp) provide a lot of smaller INDELs and several AIM-INDEL panels were originated from these sets. Some AIM-INDEL panels developed include 48 INDELs in three multiplexes (Santos *et*

al. 2010), 48 INDELs in one multiplex (Pereira *et al.* 2012), and 21 INDELs in one multiplex (Zaumsegel *et al.* 2013). Although AIM-INDELs have not as much informativeness as AIM-SNPs, but have the advantage that CE detection is used to detect polymorphic alleles, which is available in routine forensic laboratories. Therefore, AIM-INDELs provide a simple tool to derive ancestry information from a single test (Pereira *et al.* 2009).

Three statistical systems of population comparison can be used for the analysis of bio-geographical ancestry: Snipper, Principal component analysis (PCA) and STRUCTURE. Each of them uses reference population data and deduces results from the comparative patterns of variation detected.

1.4 Forensic DNA Phenotyping

When DNA profile obtained from forensic evidence does not get a match with a known suspect or in a database, and then the DNA recovered from the scene of a crime cannot help with its detection. Alternatively, DNA mass screening can be performed, but this is an expensive approach.

Another way is to derive the information from forensic evidence itself. Attempts have been made to predict the appearance of the donor of unknown stain from the evidence itself (Sundquist 2010). Most of the forensic DNA phenotyping revolved around pigmentation i.e., variation in the colouration of the human iris, head hair and skin. Research is being carried out on other traits also like body height, baldness, age, and face and hair structure (Kayser 2015).

MC1R gene was used by Forensic Science Service to predict red hair colour (Grimes *et al.* 2001). SLC24A5 gene on chromosome 15 (SNP) was identified in 2005 which impacts pigmentation (Lamason *et al.* 2005). Recently, IrisPlex assay with 6 SNPs has been developed to predict blue versus brown eye color (Walsh *et al*. 2011; Walsh *et al*. 2012). Since then a number of assay are being developed and tested to evaluate external visible characters.

Although very useful for forensic investigators and much research has been done in the recent time, but still forensic DNA phenotyping cannot be implemented in regular forensic case work. The major limitations in this regard include artificially altered appearance, multiplex genotyping issues, ethical and legal issues (Kayser 2015). It is also very limited in those populations which have less variation in pigmentation, for example, when the majority of the population has brown eyes and black hair.

1.5 DNA Degradation

Degradation of DNA is usually caused by nuclease enzymes i.e. DNAase (Paabo *et al*. 2004) .Nucleases activity results in fragmentation of DNA into smaller components. When cell membrane breaks down and release fluids, then bacteria and fungi grow quickly unless the material dries out. Most micro-organisms are capable of breaking down DNA due to the presence of nuclease enzymes (Lindahl 1993; Hofreiter *et al.* 2001). Environmental factors such as hot temperature and high humidity accelerate degradation of DNA (Robins and Furey 2001) while low temperature preserves DNA for longer period (Willerslev *et al*. 2004). Chemical oxidation is another form of degradation which involves oxidative attack on carbon- carbon double bonds of pyrimidines and purines bases of DNA causing break down of ring and base modification which results in blocking of replication and hinder the process of amplification (Moreira *et al*. 1998). However, this is not particularly important in a forensic context.

In order to perform PCR amplification, DNA template should be intact so that primers can bind and extension can happen but in case of degraded DNA, break in DNA template will stop the extension process and PCR will not be successful (Figure 1.2). The more the degraded DNA (exposed longer to environmental insults), the more will be breaks in amplification process and less PCR products will be formed (Walsh *et al.* 1992).

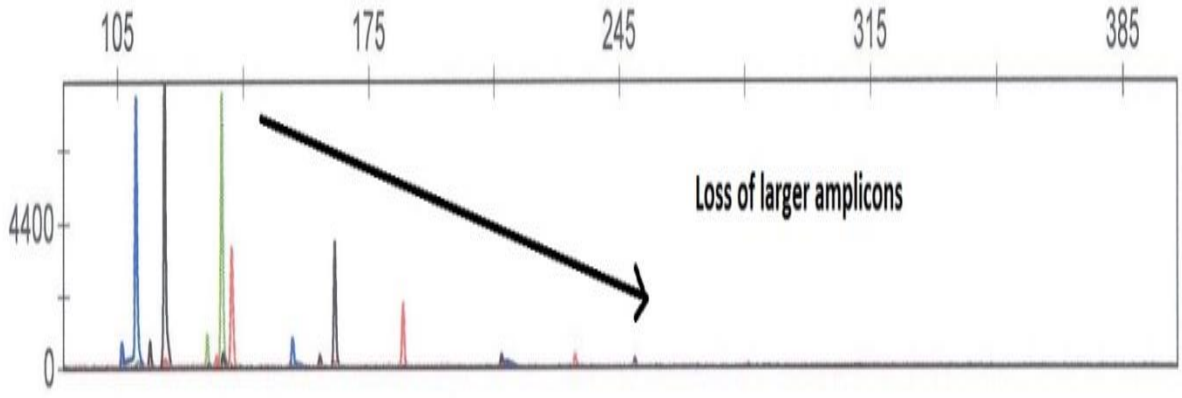

Figure.1.2 Diagram showing the profile of the loss of larger amplicons due to DNA degradation.

1.5.1 Ways to address DNA Degradation

1.5.1.1 Use of Smaller Amplicons

Use of reduced sized PCR products or amplicons could help degraded DNA to amplify successfully (Wiegand and Kleiber 2001). The reduced sized PCR products (mini amplicons) can, when the sequence environment allows, be developed from larger sized markers by re-designing their primers closer to repeat motif in order to get required reduced size (Krenke *et al*. 2002; Hill *et al*. 2008). A major problem encountered while designing mini primers was that only few markers can be put together to get into a single multiplex to accommodate into smaller size. This problem was solved by labelling primers with more fluorescent dyes (Asari *et al*. 2015). Now in the latest generation of capillary electrophoresis, it is possible to label primers with six fluorescent dyes. The best examples of 6-dye channels are GlobalFiler® and Y-Filer® Plus kits (Applied Biosystems™) in which 24 and 27 loci are accommodated in 6-dye chemistry respectively. Another tool to get the desired size of a particular marker is by adding small tail of bases, for example, ATCCGG, which helps to expand the size range of the PCR products when using CE, to the reverse primer of the marker (Ballard *et al*. 2002).

The best example of reduced sized PCR products is MiniFiler® kit (Applied Biosystems™) in which 8 STR markers from the Identifiler® kit were re-designed as mini-STRs (Mulero *et al*. 2008; Alenizi *et al.* 2009). While designing mini-STRs of MiniFiler® kit, new primers were used, so a concordance was done between two kits and 99.7 % alleles were concordant (Hill *et al*. 2007). The MiniFiler® kit was applied to a number of degraded samples and low level DNA and was found very effective in recovery of genetic materials (Luce *et al*. 2009; Muller *et al*. 2010). Another set of commercial kits containing mini-STRs are PowerPlex® ESX and PowerPlex® ESI (Promega™) which have been used to type the challenged samples (degraded and inhibited).

The advantages of mini-STRs over conventional STRs include typing of enzyme-digested DNA (Chung *et al.* 2004), burned and damaged bone samples in mass disasters (Schumm *et al*. 2004; Marjanovic *et al.* 2009) and telogen hair shafts (Muller *et al*. 2007). In another comparative study, mini-STRs performance was found better than those of conventional STRs and SNPs (Opel *et al*. 2003; Dixon *et al.* 2006).

1.5.1.2 DNA repair

Degradation of DNA might results in breaking down of double strands of DNA or contains intact DNA with single stranded nicks and DNA lesions (Lehmann 2003). Attempts have been made to repair DNA to recover genetic information during forensic DNA analysis.

For example, an enzyme "cocktail" introduced by New England Biolabs (Ipswich, MA), claimed to repair DNA damage during DNA degradation is under validation.

Taq polymerase has been an integral part of PCR amplification in forensic DNA testing. A major shortcoming of *Taq* polymerase is its inability to amplify degraded or/and inhibited DNA samples. A new type of DNA polymerase named as Y-family DNA polymerase has been developed which is found capable of successfully amplifying the damaged DNA either alone or in combination with *Taq* polymerase (McDonald *et al*. 2006).

Another treatment of damaged DNA samples was suggested by using a mixture of DNA polymerase and ligase. This repair method can be incorporated as an additional step before PCR and can increase the chances of recovering the lost loci. The method has no adverse effect on un-damaged samples (Nelson 2009).

1.5.1.3 Study of protected regions of DNA

Current research is going on to study nucleosome protected regions in the way to find which degree of damage might be done during DNA degradation. According to structural studies on nucleosome, DNA are arranged in loose manner in euchromatin and tightly packed in heterochromatin (Tan *et al.* 2011). On the basis of new findings, new PCR assays can be designed which could give better success rate with degraded DNA samples (Ioshikhes *et al.* 2006; Radwan *et al.* 2008).

1.6 PCR Inhibition

PCR inhibitions usually takes place when DNA samples are co-extracted with substances that hinders the process of amplification (Hudlow *et al*. 2011) which results in partial or complete loss of loci during amplification process. PCR inhibitors can interact with DNA directly or interfere with the DNA polymerases during PCR reaction (Mulero *et al*. 2008; Larkin *et al*. 1999).

Inhibitors can affect cell lysis during extraction of DNA and can inhibit polymerase activity through competing with magnesium ions (Akane *et al.* 1994; Wilson 1997; Radstrom *et al.* 2004). Some common PCR inhibitors and their sources are described in Table 1.2.

The effects of PCR inhibition can be reduced by diluting inhibited DNA sample and then re-amplifying it, adding more DNA polymerase (Al-Soud and Radstorm 1998), by using additive like bovine serum albumin (Comey *et al*. 1994), betaine (Al-Soud and Radstrom 2001), modifyinf the DNA extraction by adding sodium hydroxide (Bourke *et al*. 1999) or aluminum ammonium sulfate (Braid *et al.* 2003) and finally, by filtering the DNA extracts through silica columns, Centricon-100 or Microcon-100 (Comey *et al.* 1994; Lee *et al*. 2010).

Table 1.2 Table describing some common PCR inhibitors and their sources.

PCR Inhibitor	source	Reference
Calcium and Collagen	Bone and teeth	Opel *et al.* 2006 Kim *et al*. 2000
Textile Dyes	Denim Jeans	Larkin *et al.* 1999
Tannic acid	Leather substrates	Schrader *et al*. 2012
Polysaccharides	Feces	Monteiro *et al.* 1997
Bile salts	Feces	Lantz *et al.* 1997
Urea	Urine	Mahony *et al.* 1998
Melanin	Hair/Tissue	Eckhart *et al*. 2000
Heme	Blood	Schrader *et al*. 2012
Humic acid	Soil	Sutlovic *et al*. 2005 Hudlow *et al.* 2011
EDTA	Essential part of TE buffer	Rossen *et al*. 1992
Ethanol	Extraction process	Wiedbrauk *et al*. 1995
Phenol	Organic method of extraction	Katcher *et al*. 1994

1.7 Project background

Qatar is a home of 63 nationalities with a total population of about 2.24 million (Qatar Statistics 2015). The population of Qatar has grown rapidly in the last ten years with most of the increase in population due to the influx of migrant workers from different parts of the world. The crime rate is also growing in relation to the growth of population. Like other forensic DNA laboratories in the world, Qatar forensic DNA laboratory also depends upon STR chemistry (Identifiler Plus® kit from Applied Biosystems™) for forensic DNA analysis. In most of the cases, DNA profiling by using STR markers generate satisfactorily results, but in some cases especially when degraded and inhibited DNA samples are encountered, no or partial DNA profiles are obtained. In order to address the issues of DNA degradation and inhibition, special measures are taken which include the modifications in extraction methods and PCR protocols. In many of the cases, these modifications in protocols help in recovering the genetic information from the challenged DNA samples, but sometimes, due to high level of degradation and inhibition, no or little success is achieved. In order to type these challenged DNA samples, there is a need to adopt some alternative approach.

This project was designed by keeping in view the issues of DNA degradation and inhibition which are commonly encountered during forensic DNA casework. To achieve this goal, another type of PCR chemistry based on INDELs, was investigated with the possibility to incorporate into the current flow of DNA casework (based on STRs). In order to incorporate INDELs into the regular forensic case work stream like STRs, the forensic efficacy of INDELs was evaluated using samples from different populations residing in Qatar.

1.8 Aims of Study

The aim of current research is to evaluate the use of INDELs as a tool for forensic genetics within the context of Qatar. In order to achieve these specific aims, following objectives have been set:

(a) To calculate the forensic efficiency of autosomal INDELs by using genetic data generated from 5 populations groups based in Qatar

(b) To compare the forensic parameters between INDELs and STRs through the analysis of their data to evaluate their respective importance in forensic case work.

(c) To assess the effectiveness of INDELs and STRs to derive ancestry information.

(d) To develop a multiplex PCR assay with reduced amplicon length that can help to address the issue of DNA degradation.

(e) To validate the mini-INDELs multiplex assessing reproducibility, sensitivity and performance with high molecular weight and challenged DNA samples.

(f) To calculate the statistical values of mini-INDEL markers and assess their potential use as an independent identification tool or in conjunction with conventional STR markers.

CHAPTER 2
GENERAL METHODS AND MATERIALS

2.1 OVERVIEW

The general methods and materials described in this Chapter were used in the research to carry out in this study. Methods employed involving specific elements of the research have been described in the relevant Chapters.

Standard laboratory procedures were used while carrying out this research. Contamination was minimized by using lab coats, disposable gloves and masks. The extraction and PCR reactions were performed alongside negative and positive controls to identify any sporadic contamination. Bench surfaces were thoroughly cleaned with 10% bleach (sodium hypochlorite) before performing any experiment. PCR reactions were performed in dedicated PCR hoods. Pre-PCR and Post-PCR reactions were carried out in separate areas. Depending upon the nature of experiments, the samples were prepared in triplicate to maintain the accuracy and precision.

2.2 Sample Collection

Buccal swabs from 500 unrelated individuals residing in Qatar were collected with informed consent. These comprised of 100 individuals identified (both self-identified and through Nationality) as Qatari, Pakistani, Sudanese, Yemeni and Tunisian.

2.3 DNA Extraction

Genomic DNA was extracted from buccal swabs by using phenol-chloroform-isoamyl alcohol (Sigma-Aldrich™, UK). This method of extracting DNA is commonly known as the organic method. Half of the buccal swab was used as substrate for the extraction of DNA. A volume of 400 µl of stain extraction buffer (1 M Tris, 1M NaCl, 0.5 M EDTA and 20% sodium dodecyl sulphate) and 15 µl of 20 mg/ml of proteinase K (Sigma-Aldrich™, UK), were added to the substrate in a 1.5 ml microcentrifuge tube. The mixture was incubated for 2 h at 56 °C with constant shaking at 70 *x g*. After incubation, each sample was briefly vortexed and centrifuged at maximum speed (20,000 *x g*) for 1 min. The substrate was transferred to a basket (Promega™, USA) which was fitted on the same 1.5 ml microcentrifuge tube. The tube was centrifuged at full speed (20,000 *x g*) for 3 min and the basket containing dried swab head was thrown away. The clear lysate was transferred to a new Maxtract High density gel tube (Qiagen™, Germany). A volume of 400 µl of phenol-chloroform-isoamyl alcohol (PCI) mixture was added to the same gel tube and vortexed for approximately 2 min until an emulsion formed. The tube was then centrifuged at maximum speed (20,000 *x g*) for 5 min and then the aqueous layer containing DNA was transferred to a new labelled 1.5 ml microcentrifuge tube fitted with Microcon® (Merck™ Millipore). The tube was capped and centrifuged at 500 x *g* for 15 min. Then the Microcon® was removed from the tube and placed onto a new microcentrifuge tube. A volume of 300 µl of TE buffer (10 mM Tris-HCl and 0.1 M EDTA) was added to the Microcon® and the tube was again centrifuged at 500 x *g* for 10 min. Then, a volume of 100 µl of TE buffer (10 mM Tris-HCl and 0.1 M EDTA pH 8) was added to the filter of Microcon® and incubate at room temperature for 5 min. Then Microcon® was removed from tube and flipped on the new labeled 1.5 ml tube. The tube containing flipped Microcon® was centrifuged at 1000 x *g* for 5 min. After that, the Microcon® was removed and discarded and 1.5 ml tube with eluted DNA was stored at -20 °C for further use.

2.4 DNA Quantification

The extracted DNA samples were quantified using Quantifiler® Human Trio kit (Applied Biosystems™) on an ABI 7500 Real time PCR System (Applied Biosystems™) according to the manufacturer's recommended protocols.

The quantification was performed using 10 µl of reaction mix, 8 µl of primer mix and 2 µl of DNA (total volume of 20 µl). Five DNA standards were utilized (50 ng, 5 ng, 0.5 ng, 0.05 ng and 0.005 ng) to run along with the samples (see Table 2.1).

Firstly, 10 µl of reaction mix and 8 µl of primer mix for each sample were mixed together and then 18 µl of the mix was loaded into each well of the MicroAmp optical 96-well reaction plate. Then, 2 µl of DNA standard dilutions and the DNA samples were loaded into corresponding wells. After loading, the optical adhesive cover was used to seal the plate. The plate was then centrifuged at 250 *x g* for 1 min using 5740 centrifuge (Eppendorf™, Germany) to remove any air bubbles.

Then plate was loaded on the ABI 7500 Real-Time PCR machine. The thermal cycler protocol was used per manufacturer's instructions: stage 1, 95 °C for 2 min for 1 cycle; stage 2 at 95 °C for 9 s followed by 60 °C for 30 s for 40 cycles. After each run, the data were analyzed and DNA concentration for each sample was estimated in ng/µl.

Table 2.1 Table showing the preparation of standards used in Applied Biosystems™ Quantifiler® Human Trio kit.

Standard	Concentration (ng/ µl)	Example Volumes	Dilution Factor
STD.1	50	10 µl (100 ng/µl of stock) + 10 µl of dilution buffer	2X
STD.2	5.000	10 µl (STD.1) + 10 µl of dilution buffer	10X
STD.3	0.500	10 µl (STD.2) + 10 µl of dilution buffer	10X
STD.4	0.050	10 µl (STD.3) + 10 µl of dilution buffer	10X
STD.5	0.005	10 µl (STD.4) + 10 µl of dilution buffer	10X

2.5 Amplification of STRs

The amplification of 15 STRs and amelogenin was performed using Applied Biosystems™ Identifiler Plus® kit using a ABI Thermal cycler 9700 (Applied Biosystems™).

2.5.1 Amplification of STRs using Identifiler Plus® Kit

The PCR mix was made using half volume for each sample by mixing 5 µl of reaction mix and 2.5 µl of primer mix. Then, a total volume of 5 µl containing nuclease-free water and template DNA was added (volumes based on quantity of DNA). Positive and negative controls were also included in each batch of samples for quality control purpose. The ABI thermal cycler 9700 was programmed using the conditions outlined in Table 2.2.

After the cycling protocol was completed, PCR products were stored at –20 °C prior to analysis.

Table 2.2 Table showing the standard PCR protocol for Applied Biosystems™ Identifiler Plus® kit (Taken from the Identifiler Plus® kit Handbook).

Temperature	Time	Number of cycles
95 °C	11 min	**1**
94 °C **59 °C**	20 s 3 min	**27 cycles**
60 °C	10 min	**1**
4 °C	**∞**	

2.6. Amplification of INDELs

The amplification of 30 biallelic INDELs and amelogenin was carried out using Qiagen™ Investigator® DIPplex kit using ABI Thermal cycler 9700 (Applied Biosystems™).

2.6.1 Amplification of INDELs using Investigator® DIPplex Kit

The master mix was made for each sample using reaction mix of 1.25µl, primer mix of 1.25 µl and 0.15 µl of Multi Taq2 DNA Polymerase. Then, the nuclease-free water and template DNA (volumes based on quantity of DNA) were added to bring final volume of 6.25 µl. Positive and negative controls were also included in each batch of samples for quality control purpose. The ABI thermal cycler 9700 was programmed using the conditions outlined in Table 2.3.

After the cycling protocol was completed, PCR products were stored at –20 °C prior to analysis.

Table 2.3 Table showing the standard PCR protocol for Qiagen™ Investigator® DIPplex kit, (Taken from the Investigator® DIPplex kit Handbook).

Temperature	Time	Number of cycles
94 °C	4 min	**1**
94 °C **61 °C** **72 °C**	30 s 120 s 75 s	**30 cycles**
68 °C	60 min	**1**
10 °C	**∞**	

2.7 Capillary Electrophoresis

DNA fragment analysis was carried out on 3130XL Prism Genetic Analysers (Applied Biosystems™) using 36 cm long capillaries and POP-4. Following parameters were used: run temperature 60 °C, injection time 10 s, injection voltage 1.6 kVs and run voltage 15 kV.

For fragment analysis, samples were prepared by adding 1 µl of PCR product to 8.5 µl of Hi-Di formamide and 0.5 µl LIZ-500 internal size standard (Applied Biosystems™). Then, the samples were denatured for 3 min at 95 °C and then placed on ice for 3 min.

2.8 Data Analysis

The data obtained from the capillary electrophoresis (CE) were analysed using GeneMapper® Software v4.1 (Applied Biosystems™) and DIP sorter software (Qiagen™). The analysis parameters used in GeneMapper® software for the data were kept the same for every experiment (Table 2.4). DIP sorter software developed by Qiagen™ was used to convert INDEL profiles into excel format which later on, was used conveniently for statistical calculations.

Table 2.4 Table showing the parameters used for the analysis of PCR fragments using GeneMapper® Software v4.1.

Parameters	Values
Analysis Range	Partial Range (2200-15000)
Baseline Window	51 pts (points)
Minimum Peak Half Width	2 pts
Peak Detection	50 RFU
Peak Window Size	15 pts
Polynomial Degree	3 pts
Size Call Range	All Sizes
Size Calling Method	Local Southern
Slope Threshold for peak start/end	0-0

2.9 Statistical Analysis

The genetic data were analysed from both markers (INDELs and STRs) to evaluate their comparative forensic importance in situations like paternity testing and individual identifications. Forensic parameters which included Discrimination Power (DP), Match Probability (MP) and Polymorphic Information Content (PIC), Typical Paternity Index (TPI) and Power of Exclusion (PE), were estimated using Powerstats V12 software (Tereba 1999). Table 2.5 showing the formulae of different forensic parameters.

Population indices which included Observed Heterozygosity (Ho), Expected Heterozygosity (He) and P-value for Hardy–Weinberg Equilibrium (HWE), F_{ST} value and estimation of linkage disequilibrium (LD) for each pair of INDEL markers in the Investigator™ DIPplex kit were calculated for each INDEL with Arlequin 3.5 software (Excoffier *et al.* 2010).

STRUCTURE 2.3.4 software (Pritchard *et al*. 2010) was used to estimate the population structure based on the STR and INDEL data. This software used a Bayesian probabilistic clustering approach to estimate the population of individuals on the basis of genetic data.

An online software called Snipper app suite v2.0 (Phillips *et al.* 2007) was also used to predict ancestral origin of individuals. This programme was based on a Bayesian system and could be used to identify ancestry and also estimate misclassification rates by testing CEPH human genome diversity cell line panel containing samples of known geographic origin.

Table 2.5 Table showing the formulae of different forensic parameters along with their explanations.

Forensic Parameter	Formuale	Explainations
Heterozygositiy	$h = \frac{n_h}{n}$	Where n_h is the number of individual observations with two alleles and n is the total number of individuals.
Probability of Match	$pM = \sum_{i=a}^{n} \sum_{j \geq 1}^{n} P_{ij}^2$	Where i and j represent the frequencies of all possible alleles a through n, P_{ij} represents the frequencies of all possible genotypes.
Typical Paternity Index	$PI_{typical} = \frac{1}{2H}$	Where H is homozygosity of a locus.
Power of Discrimination	$P_d = 1 - pM$	Where pM is probability of match.
Power of Exclusion	$PE = h^2(1-2hH^2)$	Where h is the heterozygosity and H is the homozygosity of a locus.

CHAPTER 3
FORENSIC EVALUATION OF AUTOSOMAL INDELS IN QATAR

3.1. OVERVIEW

INDEL polymorphisms have been studied in different parts of the world, but the Gulf region in general and Qatar in particular have no data relating to its different resident populations. In this Chapter, INDEL data have been generated from 500 individual's samples from the five population groups (Qatari, Pakistani, Sudanese, Yemeni and Tunisian) based in Qatar using the Investigator® DIPplex kit (Qiagen™). At the same time, 15 STRs (using Applied Biosystems™ Identifiler Plus® kit) have also been genotyped for the same 500 DNA samples (which have been used for INDELs analysis) to compare the forensic efficiency with that of INDELs.

The generated data were statistically analyzed for both systems (INDELs and STRs) to evaluate their comparative forensic importance in situations like paternity testing and individual identifications using the parameters Discrimination Power (DP), Match of Probability (MP) and Polymorphic Information Content (PIC), Typical Paternity Index (TPI) and Power of Exclusion (PE), Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD).

3.1.1. Objectives and Aims

The aim of the current research was to evaluate the use of INDELs as a tool for forensic genetics within the context of Qatar. To achieve this, following objectives were set.

(a) To optimise and validate the INDEL system for processing of crime scene and reference samples.

(b) To generate and evaluate the population genetic data of 30 autosomal INDELs (using QIAGEN™ Investigator® DIPplex kit) using different population groups based in Qatar.

(c) To generate and evaluate the population genetic data of 15 autosomal STRs (using Applied Biosystems™ Identifiler® plus kit) using different population groups based in Qatar.

(d) To compare INDELs and STRs to evaluate their respective importance in forensic case work.

3.2. Materials and Methods

The methods and statistical analysis used in this Chapter was already described in Chapter2.

3.3. RESULTS

3.3.1 Optimisation of Investigator® DIPplex Kit

INDELs analysis using Investigator® DIPplex kit (Qiagen™) was optimised in terms of quantity of input genomic DNA (sensitivity) and reaction volume of PCR mix.

3.3.1.1 Optimisation of Input DNA (Sensitivity study)

The sensitivity study for DIPplex kit was designed to determine the minimum amount of DNA at which samples could generate a good quality profile with sufficient peak height ratio and well-balanced peaks. Three different amounts of genomic DNA (9947A) were tested (0.1 ng, 0.3 ng and 0.5 ng).

The sensitivity results of DIPplex kit indicated that its loci were capable of amplifying efficiently for 0.5 ng with good peak heights and quality profiles, but drop-out of alleles were observed for each of the samples 0.1 ng and 0.3 ng (Table 3.1 and Figure 3.1).

Table 3.1 Table showing the sensitivity results of DIPplex kit for different quantity of input DNA including the number of drop outs from total number of alleles and average peak heights(n = 3).

DNA concentration (ng)	# of alleles drop-out (out of 14 alleles in blue dye)	Average Peak Heights (RFU)	Std. Deviation for peak heights (RFU)
0.1	2	900	150
0.3	2	1500	300
0.5	0	2500	100

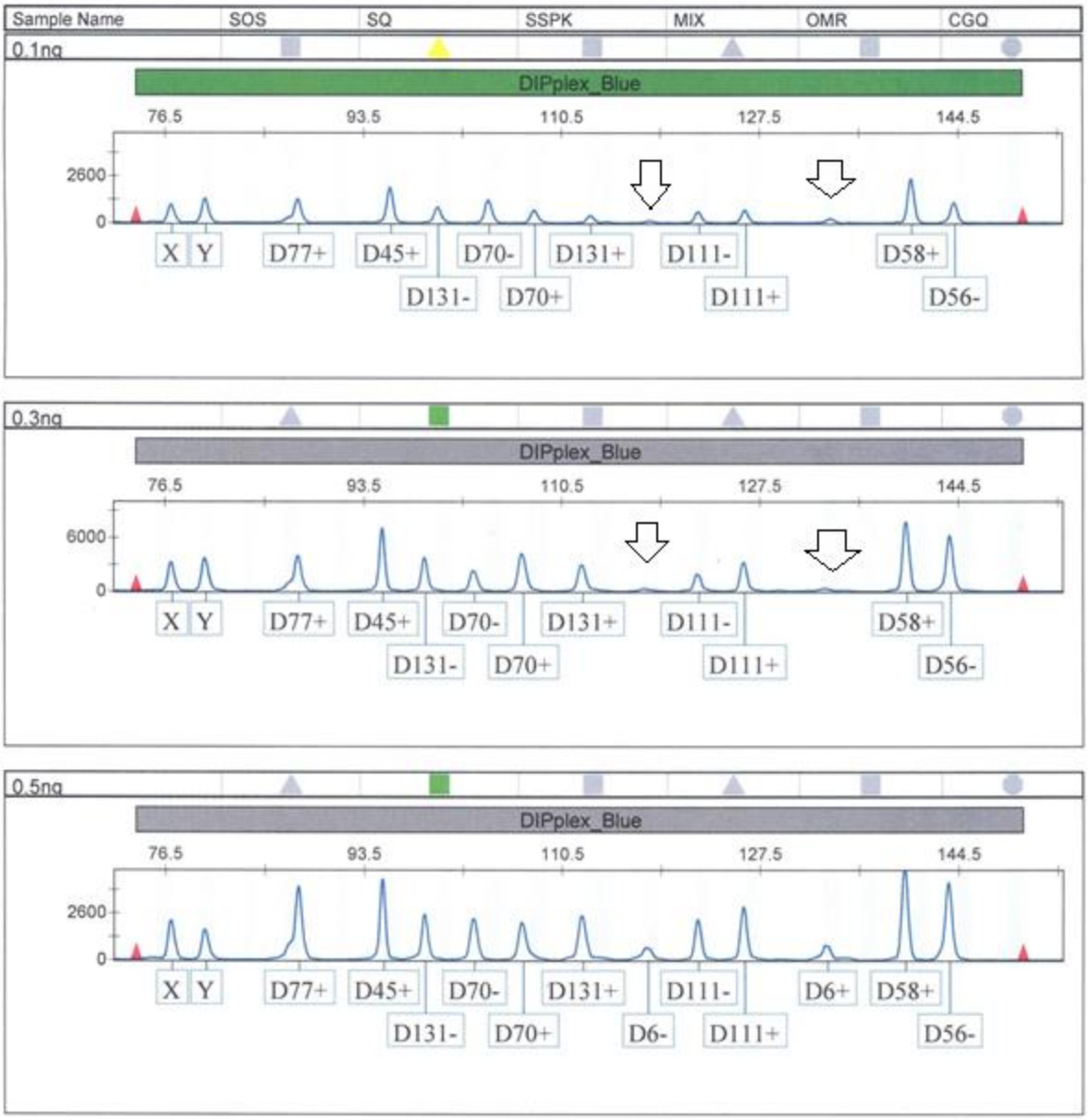

Figure 3.1 Original traces showing the electropherogram of the Qiagen™ Investigator® DIPplex Kit using 0.1 ng, 0.3 ng and 0.5 ng Control DNA 9947A (arrows indicate the drop outs of alleles); n = 3

3.3.1.2 Optimisation of reaction volume

To determine the optimum reaction volume of PCR mix, a set of volumes of one half, one quarter and one fifth of the recommended 25 µl was tested using 0.5 ng of control DNA (9947A). The results showed that all of three reaction volumes were capable of amplifying all of the loci efficiently. The volume of one half and one quarter were better than that of one fifth in terms of peak balance, peak morphology and peak height (Table 3.2 and Figure 3.2).

Table 3.2 Table showing the sensitivity results of Qiagen™ Investigator® DIPplex kit for different quantity of input DNA including the number of drop-outs from total number of alleles and average peak heights (n = 3).

PCR Reaction Volume	# of alleles drop-out (out of 14 alleles in blue dye)	Average Peak Heights (RFU)	Std. Deviation for peak heights (RFU)
One Half	0	1000	200
One Quarter	0	1800	150
One Fifth	0	700	200

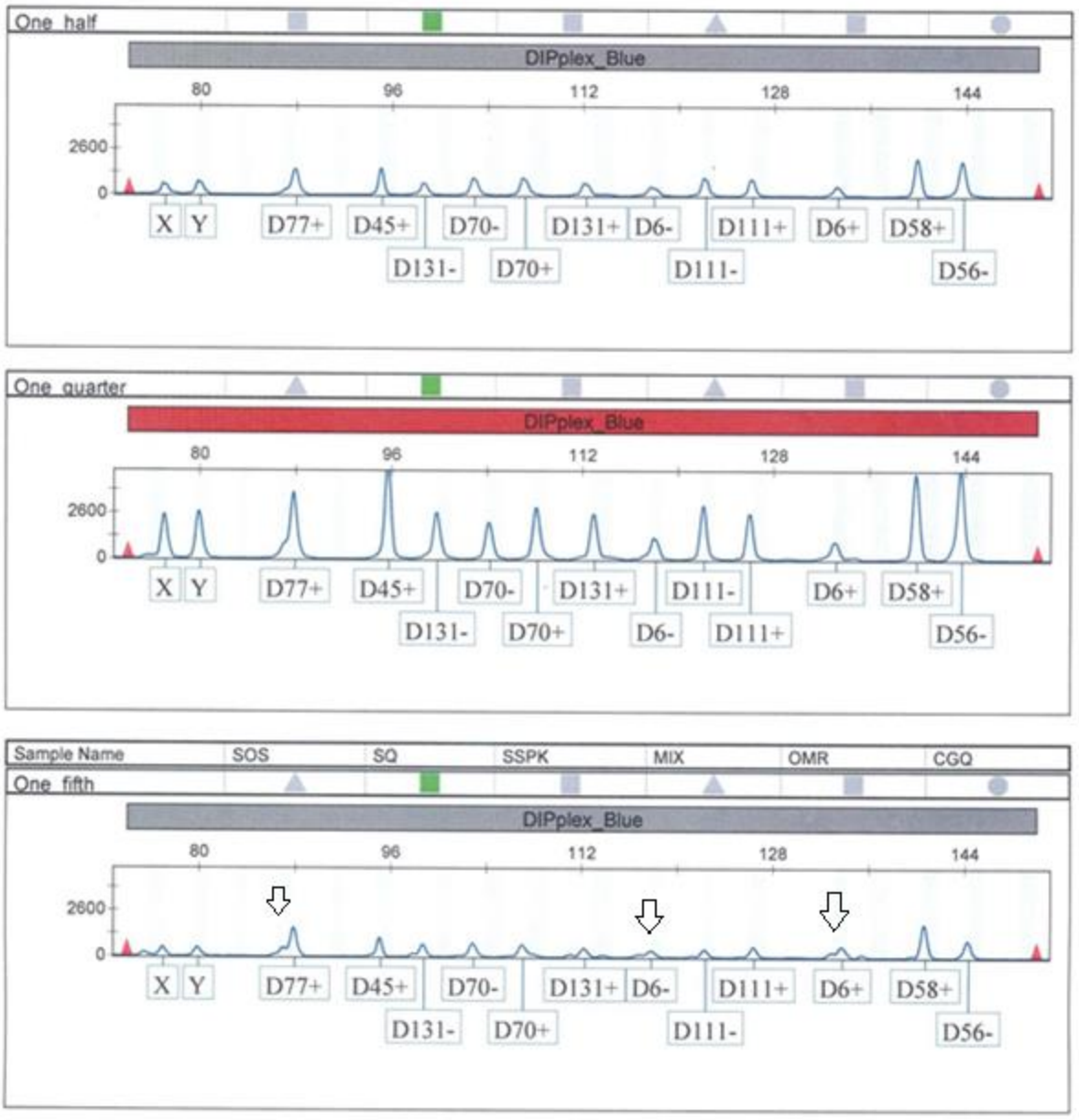

Figure 3.2 Original traces indicating the electropherogram of 0.5 ng of control DNA 9947A amplified using one half, one quarter and one fifth volume of PCR mix with Qiagen™ Investigator® DIPplex kit (arrows indicate the poor peak morphology); n = 3

3.3.2 Quantification of DNA samples

DNA samples extracted from 500 individuals were quantified using Quantifiler® Human Trio kit (Applied Biosystems™). While quantifying the samples, two sets of DNA standards were run to get accuracy in results (Figure 3.3).

Quantifiler® Human Trio kit provides information about the quality and quantity of DNA in four ways i.e. IPC (internal positive control) which can give an indication of inhibition and large autosomal target. In turn, this reveals the presence of quantity of DNA for larger amplicons, the small autosomal target indicated the quantity of DNA for smaller amplicons and Y-chromosome target quantified male DNA. To predict the degradation in a DNA sample, degradation index can be derived by dividing amounts of large autosomal to small autosomal. Table 3.3 shows the quantification results (ng/µl) of five Qatari DNA samples (from the total elution volume of 100 µl).

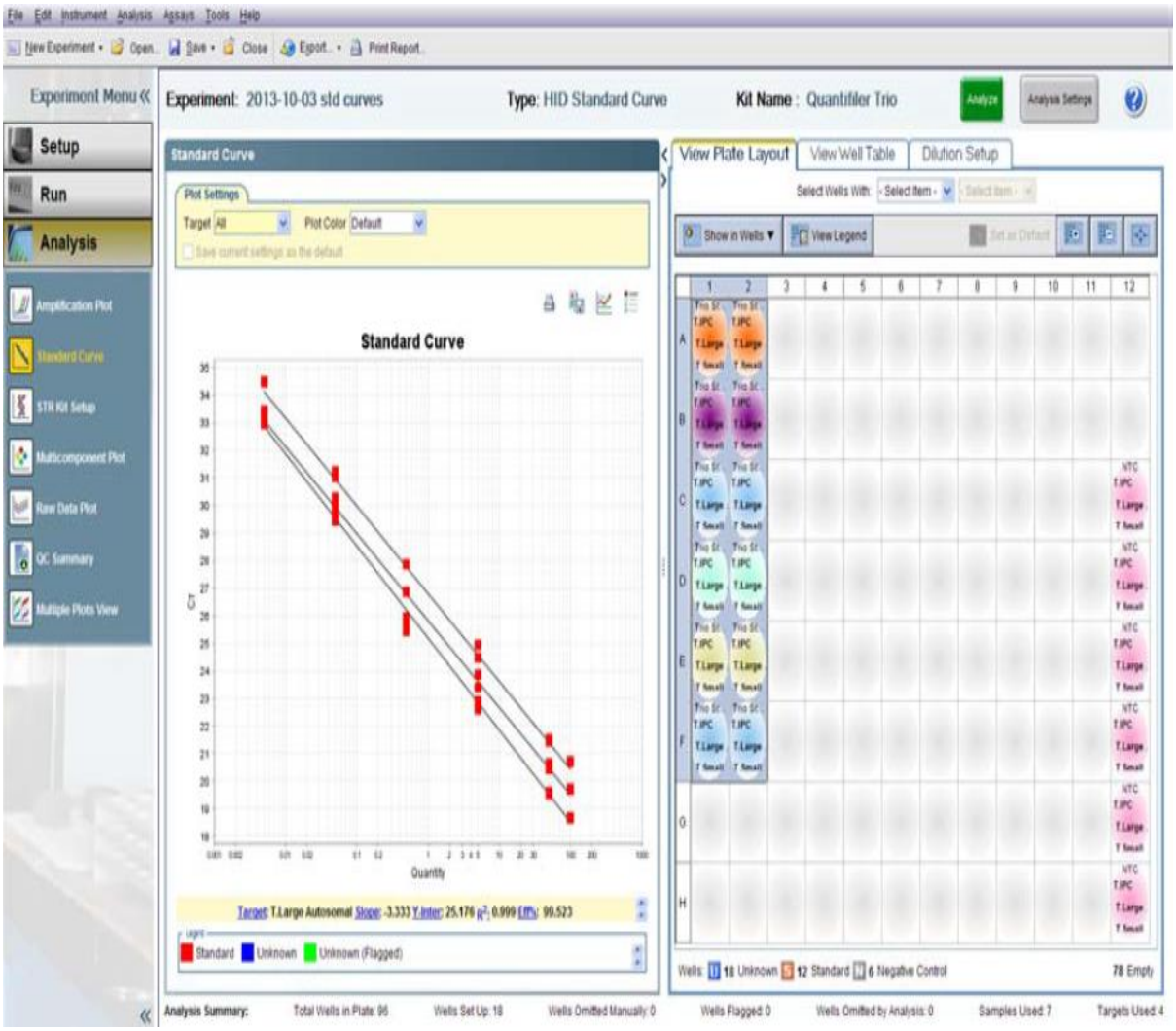

Figure 3.3 Data showing an example of two standard curve plots, each containing small Autosomal, large Autosomal, and male CT values for five standards.

Table.3.3 Table showing the quantitation results of five Qatari samples using Applied Biosystems™ Quantifiler® Human Trio kit.

Results Table

Well	Sample	Target	Task	Quantity	CT
F1	QAT-001	T.IPC	Unk		27.7808
F1	QAT-001	T.Large Autosomal	Unk	0.1538	28.3697
F1	QAT-001	T.Small Autosomal	Unk	0.1682	30.3458
F1	QAT-001	T.Y	Unk	0.1078	29.6713
G1	QAT-002	T.IPC	Unk		28.1431
G1	QAT-002	T.Large Autosomal	Unk	0.2087	27.9158
G1	QAT-002	T.Small Autosomal	Unk	0.215	29.9957
G1	QAT-002	T.Y	Unk	0.1717	28.9959
H1	QAT-003	T.IPC	Unk		27.9665
H1	QAT-003	T.Large Autosomal	Unk	0.0018	34.9834
H1	QAT-003	T.Small Autosomal	Unk	0.0228	33.1985
H1	QAT-003	T.Y	Unk	0.0214	32.0222
A2	QAT-004	T.IPC	Unk		27.8688
A2	QAT-004	T.Large Autosomal	Unk	0.0197	31.4199
A2	QAT-004	T.Small Autosomal	Unk	0.0655	31.6925
A2	QAT-004	T.Y	Unk	0.0877	29.9715
B2	QAT-005	T.IPC	Unk		27.37
B2	QAT-005	T.Large Autosomal	Unk	0.1491	28.4155
B2	QAT-005	T.Small Autosomal	Unk	0.3798	29.1835
B2	QAT-005	T.Y	Unk	0.3774	27.8523

3.3.3 DNA profiling of 500 samples

A total of 500 DNA samples from individuals belonging to five nationalities based in Qatar were amplified using Investigator® DIPplex kit (Qiagen™) and Identifiler Plus® kit (Applied Biosystems™). ABI Thermal cycler 9700 (Applied Biosystems™) was used to perform the amplification process for INDELs and STRs using the Manufacturers' recommended protocols, apart from the reduced volume.

The results of the DNA profiling showed that 95% of the samples were amplified successfully in their first batches using 0.5 ng DNA templates with Investigator® DIPplex and Identifiler Plus® kits. In order to achieve 0.5 ng of target DNA concentration, required volume of eluted DNA was added directly from the stock of eluted DNA. For example, if DNA concentration was 0.1 ng, then a volume of 5 µl was used. In the cases, where samples contained high quantity of DNA, required dilutions of samples were made. For example, if DNA concentration was 5 ng, then a dilution of ten was done by diluting DNA sample 10 times to achieve 0.5 ng concentration. For a few samples, where quantification of DNA was very low, the samples were re-extracted with the remaining substrate using an extended incubation time (for 3-4 h) and decreased elution volume (50 µl). These re-extracted samples were re-quantified to confirm that the required amount of DNA had been achieved for amplification process.

For each batch of amplification for INDELs and STRs (i.e. 90 samples), 3 allelic ladders were run along with the samples to get consistent and precise results. Negative and positive controls were also run along with each batch of samples to detect any contaminations.

The amplified DNA from the samples was detected on Applied Biosystems™ Genetic Analysers 310, 3130XL and 3500 according to recommended protocols in different batches. GeneMapper® Software v4.1 (Applied Biosystems™) was used to analyse the data generated during capillary electrophoresis. DIP sorter software developed by Qiagen™ was used to convert INDEL profiles into excel format which was used for statistical calculations. Most of the DNA profiles generated from INDELs and STRs had good quality peaks with good morphology and reasonable heights. In some profiles, pull ups, split peaks and other artefacts were observed, and then these poor quality samples were re-injected with reduced amount of PCR

products (0.5 µl) to achieve quality profiles. Some profiles generated during the analysis of INDELs and STRs are shown in Figures 3.4-3.11.

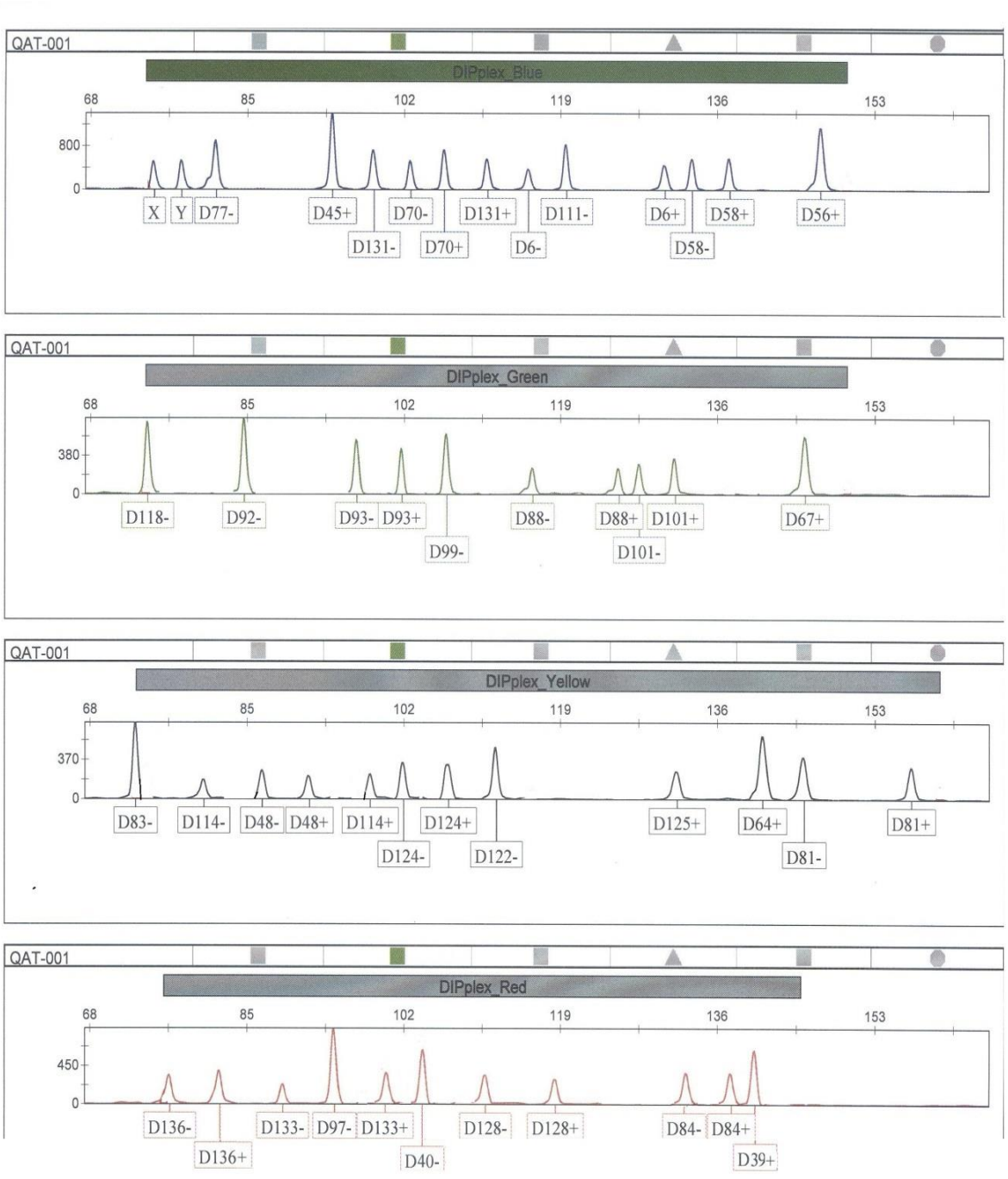

Figure 3.4 Original trace indicating a DNA profile generated from a Qatari sample using Qiagen™ Investigator® DIPplex kit.

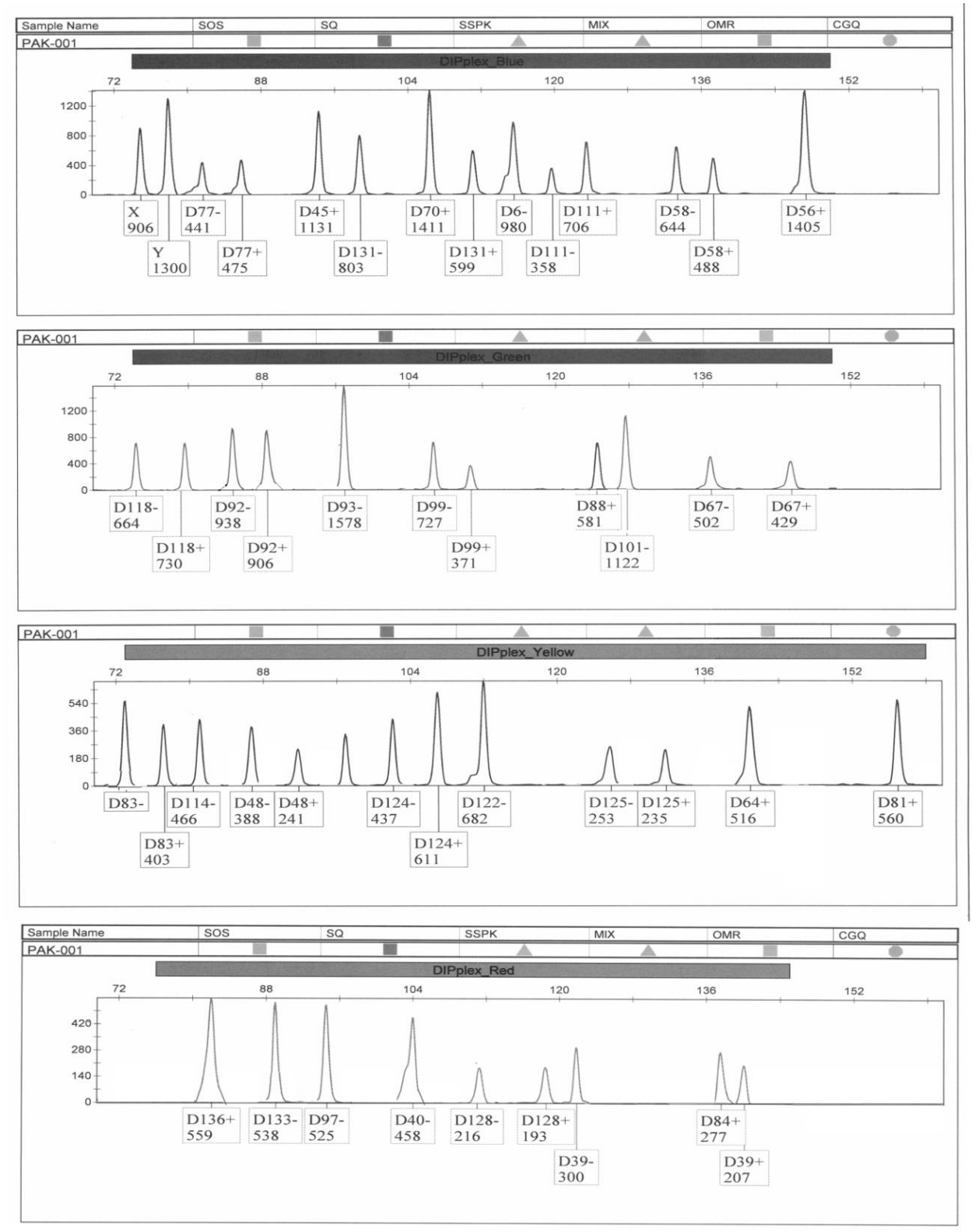

Figure 3.5 Original trace indicating a DNA profile generated from a Pakistani sample using Qiagen™ Investigator® DIPplex kit.

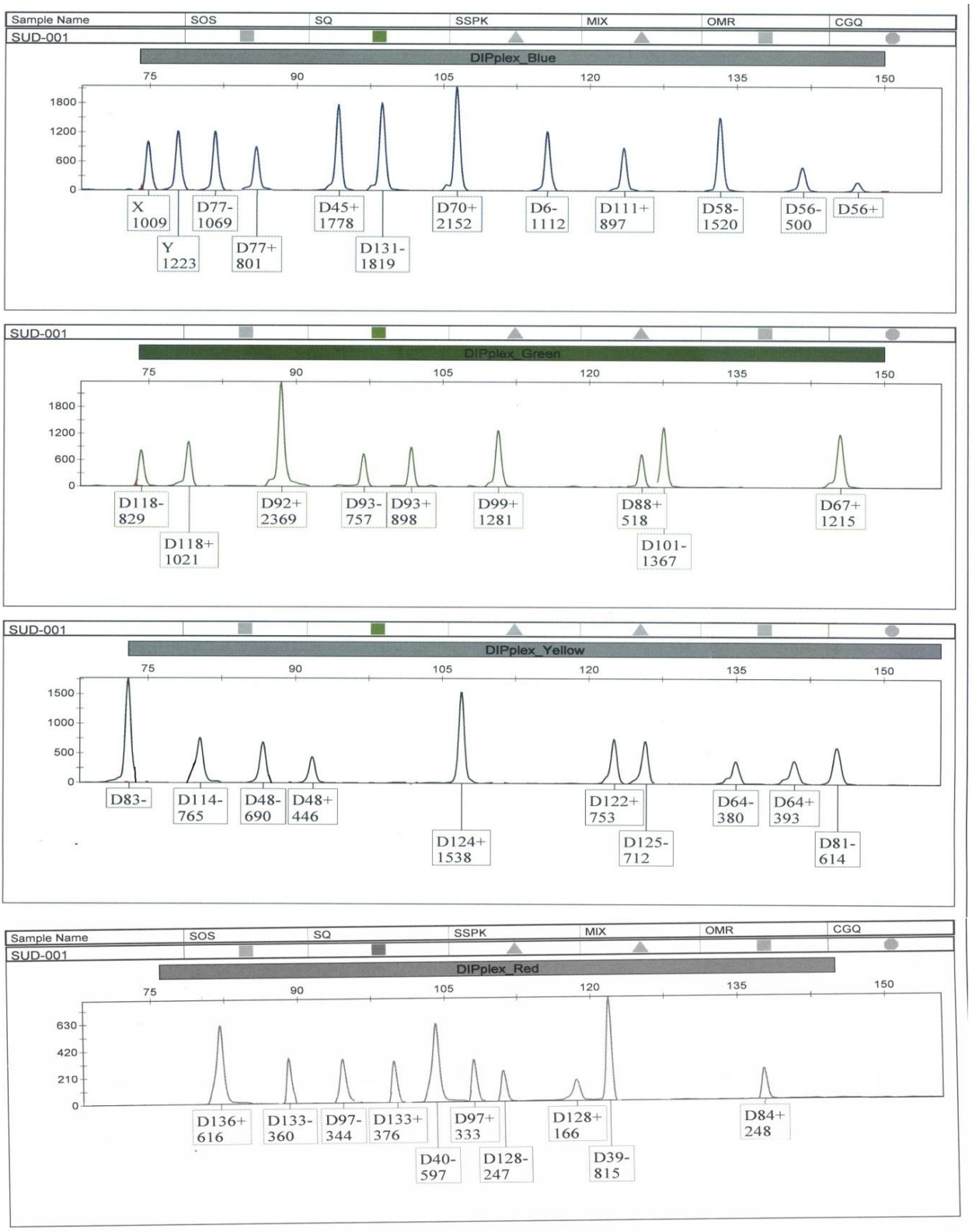

Figure 3.6 Original trace indicating a DNA profile generated from a Sudanese sample using Qiagen™ Investigator® DIPplex kit.

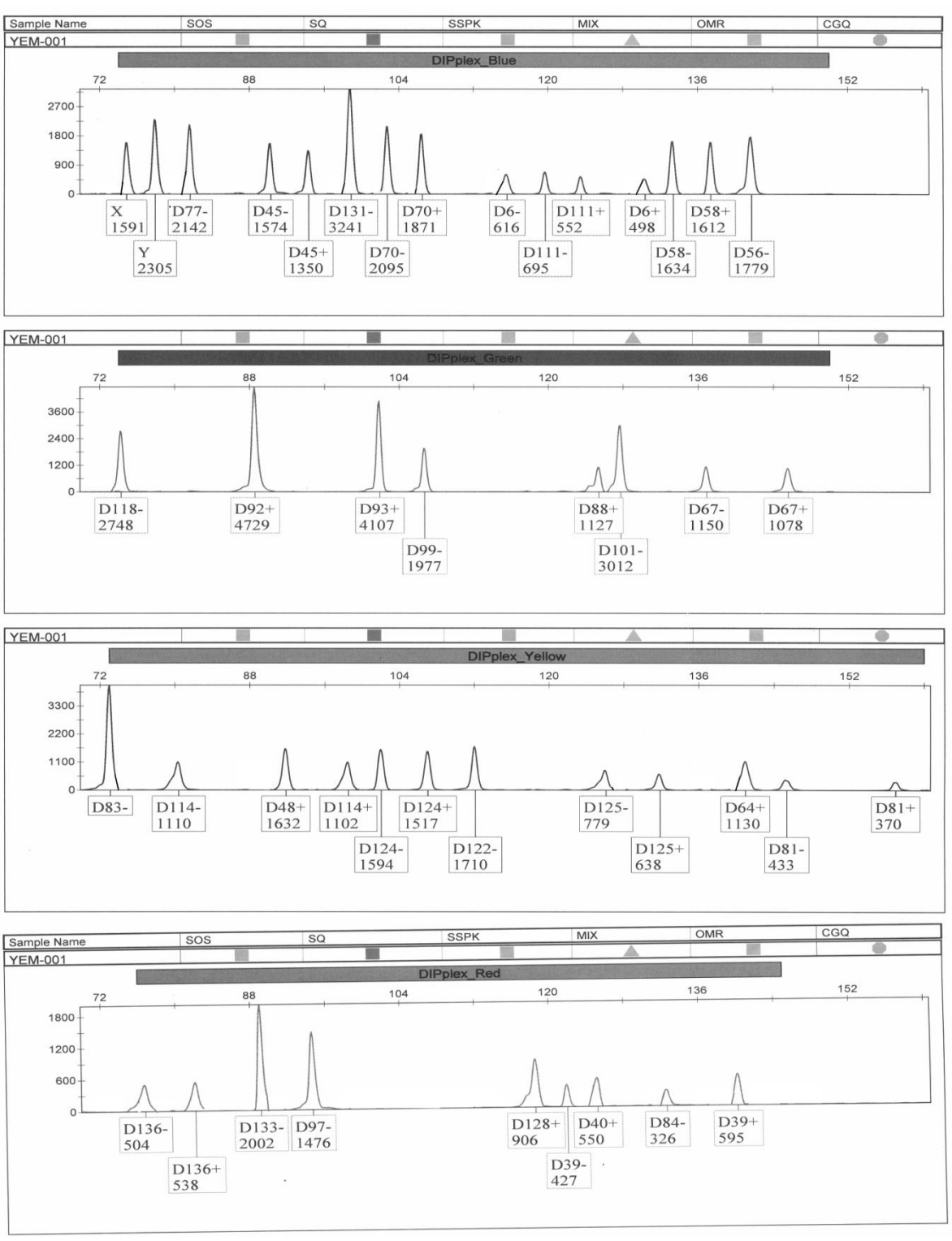

Figure 3.7 Original trace indicating a DNA profile generated from a Yemeni sample using Qiagen™ Investigator® DIPplex kit.

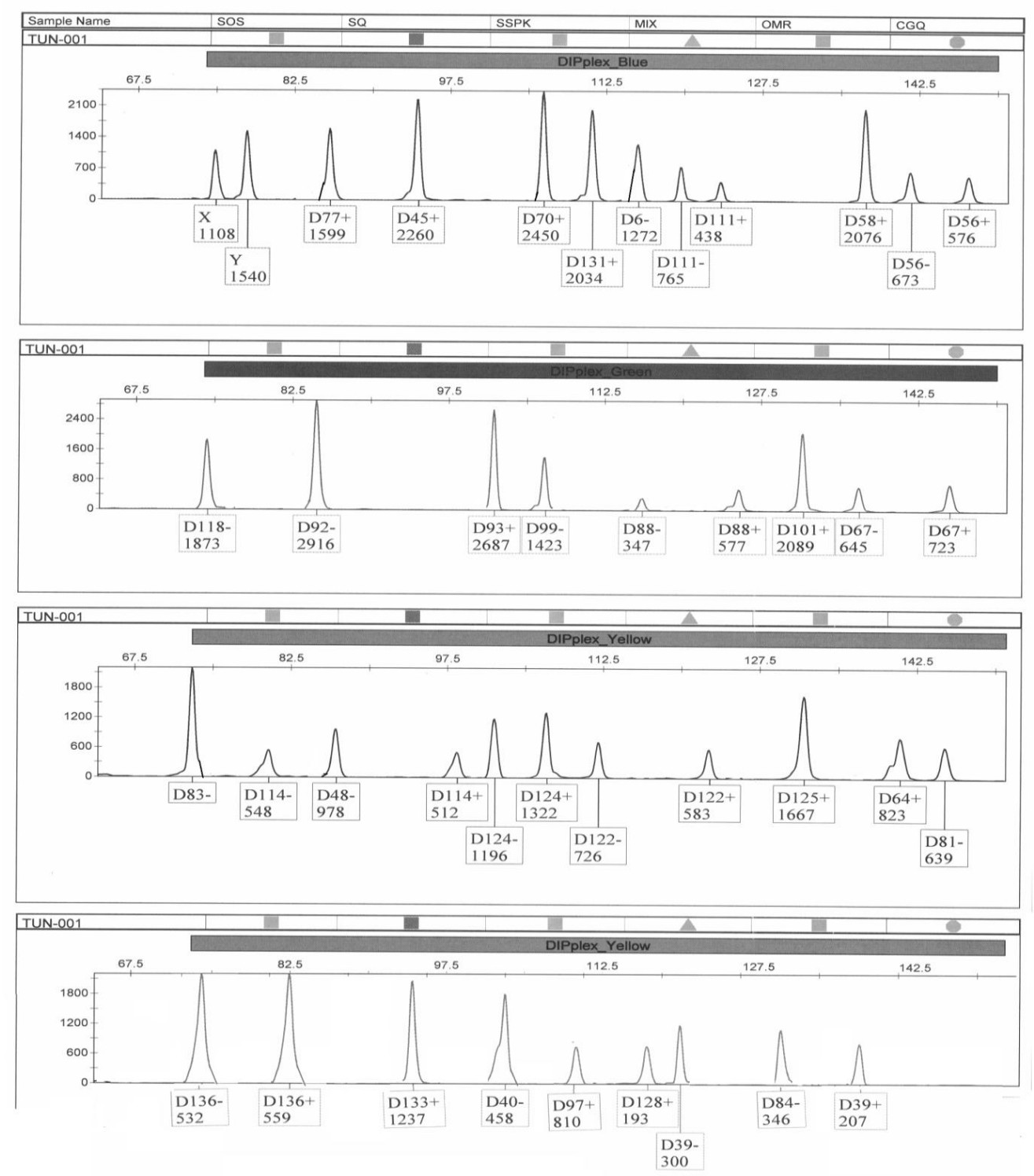

Figure 3.8 Original trace indicating a DNA profile generated from a Tunisian sample using with Qiagen™ Investigator® DIPplex kit.

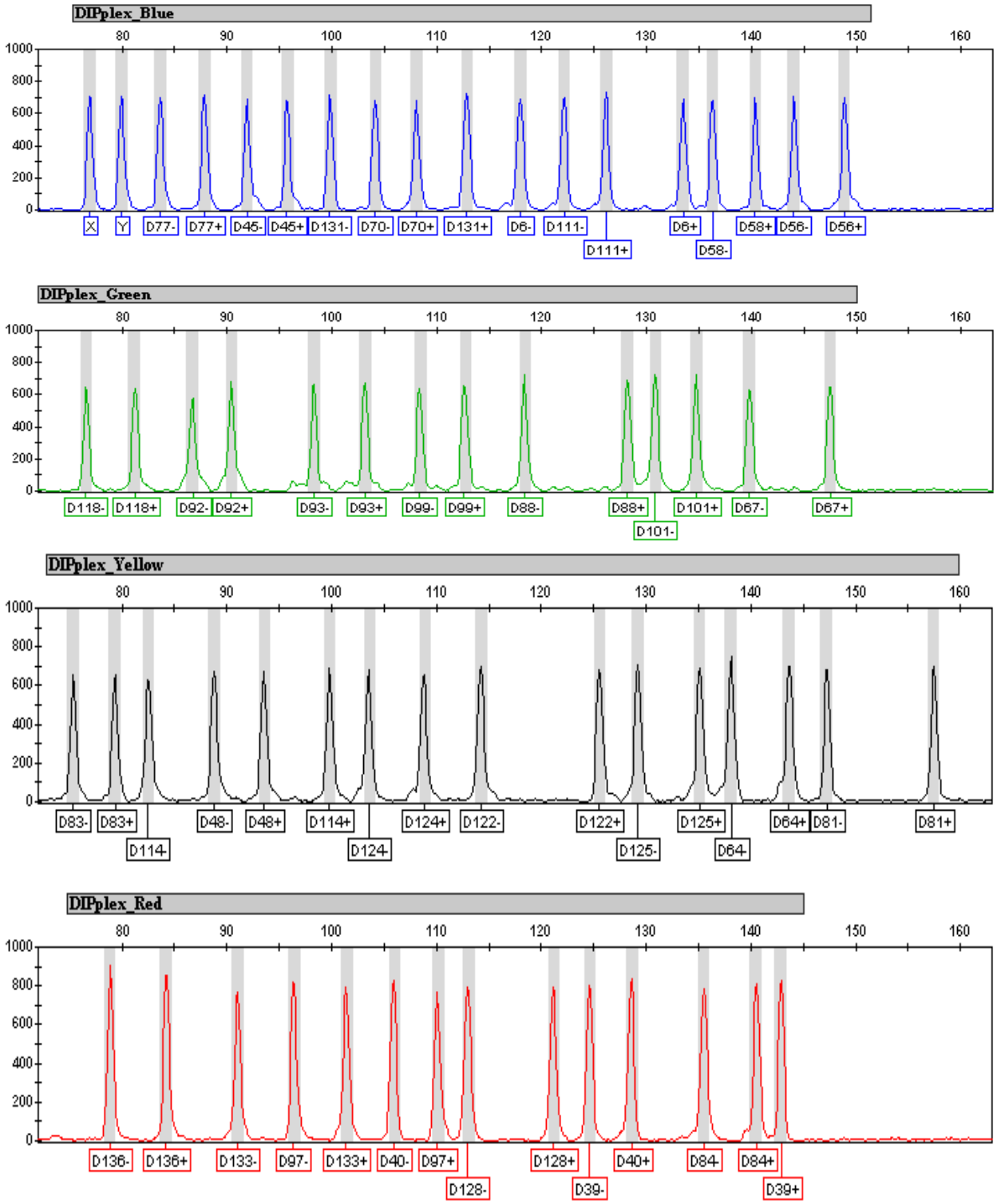

Figure 3.9 Original trace indicating the electropherogram of the allelic ladder of Qiagen™ Investigator® DIPplex kit containing all the alleles.

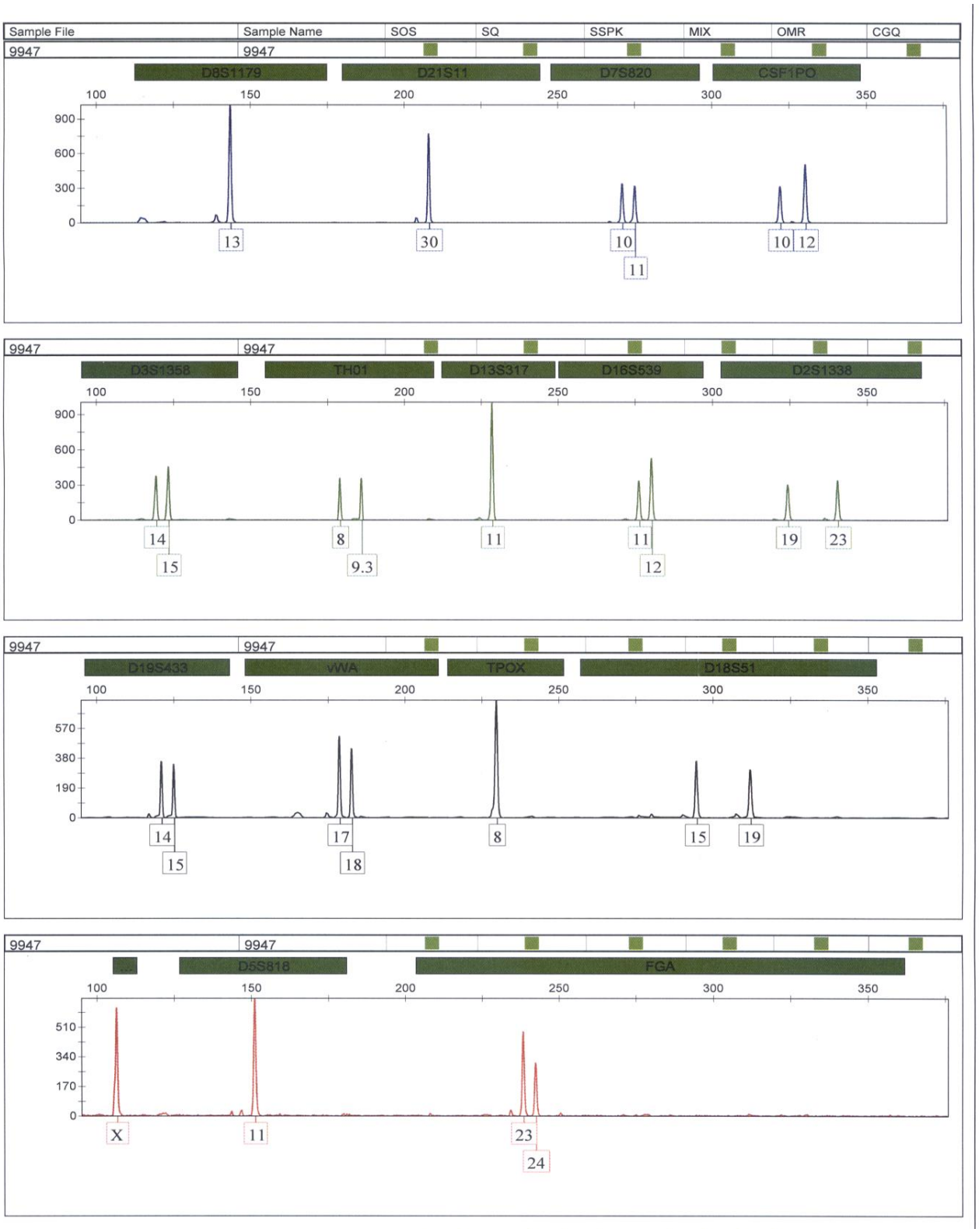

Figure 3.10 Original trace indicating the DNA profile generated from control DNA 9947A using Applied Biosystems™ Identifiler Plus® kit.

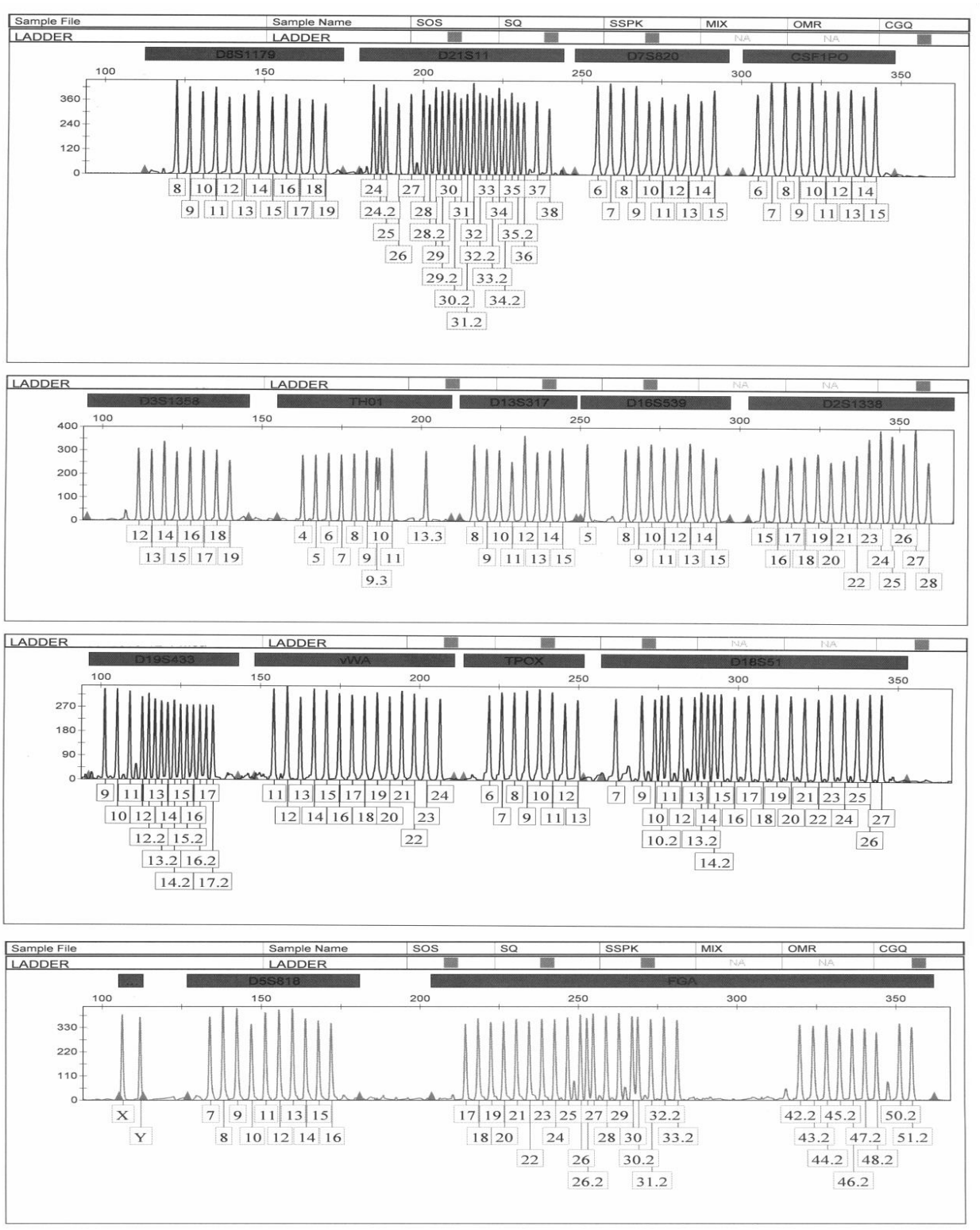

Figure 3.11 Original trace indicating the Electropherogram of the allelic ladder of Applied Biosystems™ Identifiler Plus® kit containing all the alleles.

3.3.4 Statistical Analysis of INDELs Data

Allele frequencies, expected and observed heterozygosities, exact test of Hardy-Weinberg equilibrium, exact tests of linkage disequilibrium and F_{ST} genetic distances were calculated for 500 INDEL profiles with Arlequin v3.5 software (Excoffier *et al.* 2010). While forensic parameters were assessed using PowerStats V12 software (Tereba 1999). The allele frequencies and forensic efficiency parameters for the 30 INDEL loci in the five population groups in Qatar are shown in Tables 3.4-3.8.

3.3.4.1 Allele frequencies

The variation in allele frequencies for five population groups is illustrated using HLD77 locus in Figure 3.12. The allele frequencies for HLD77- for five populations ranged from 0.61 (Tunisian) to 0.73 (Sudanese).

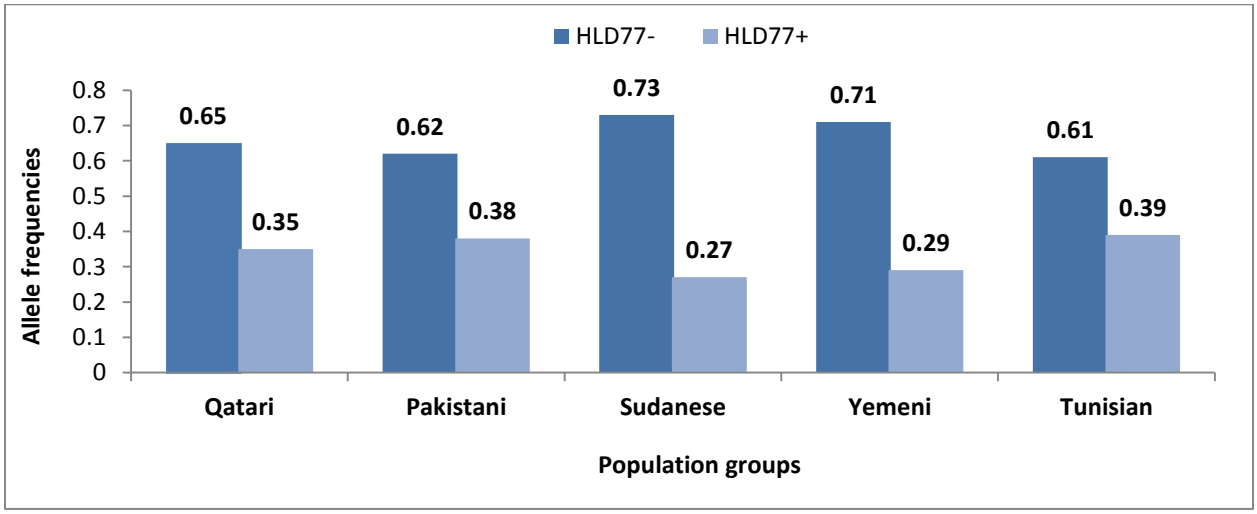

Figure 3.12 Bar diagram showing allele frequencies for deletion and insertion alleles of HLD77 locus.

3.3.4.2 HWE

While calculating the *p*- value for Hardy-Weinberg equilibrium, 4 (Qataris), 7 (Pakistanis), 5 (Sudanese), 8 (Yemenis) and 6 (Tunisians) loci showed departure ($P < 0.05$). But, after applying Bonferroni's correction at $P < 0.00166$ (0.05/30), there was no deviation from Hardy-Weinberg equilibrium. The expected heterozygosities ranged from 0.290 (Sudanese) to 0.500 (Tunisian) with a mean value of 0.395, while the observed heterozygosities ranged from 0.250 (Sudanese) to 0.770 (Pakistani) with a mean value of 0.510. Figure 3.13 showed the comparative values of observed and expected heterozygosities of one locus as an illustration.

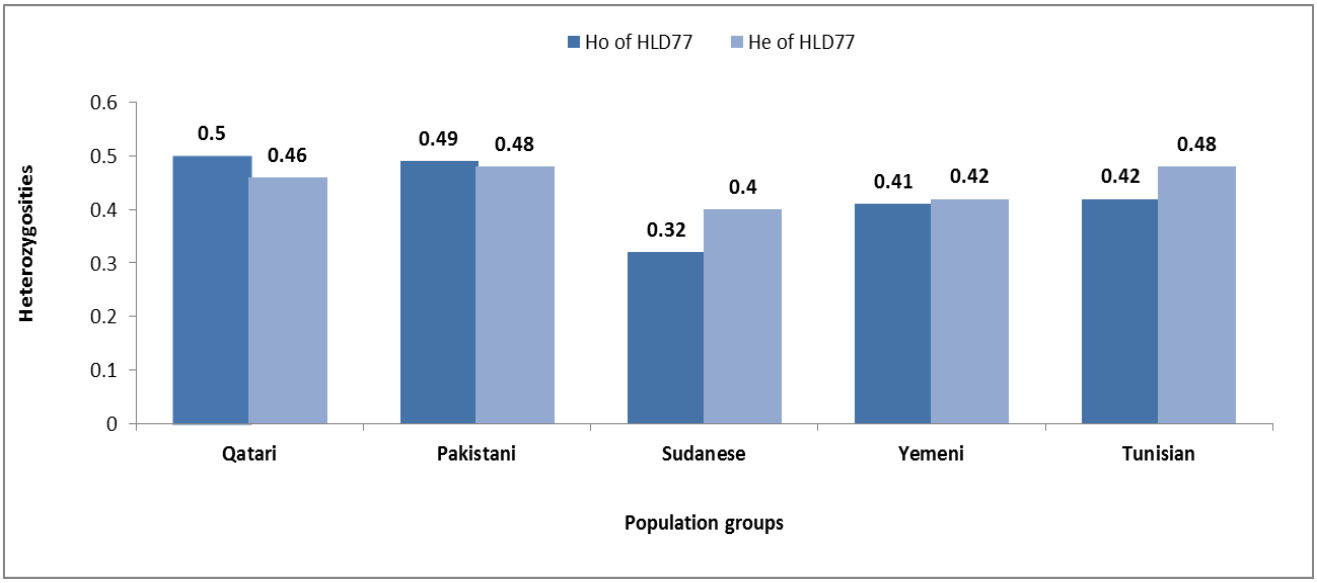

Figure 3.13 Showing the values of observed and expected heterozygosities for HLD77 locus.

3.3.4.3 Linkage disequilibrium

The pairwise linkage disequilibrium (LD) between 30 INDEL loci in each population group of Qatar was also tested (Table 3.9). The results revealed that 32 pairs in Qataris, 34 pairs in Pakistanis, 30 pairs in Sudanese, 23 pairs in Tunisians and 36 pairs in Yemenis were detected demonstrating significant linkage disequilibrium ($P < 0.05$) but after applying Bonferroni's correction, there was no significant linkage disequilibrium among INDELs on the same chromosome. So, the assumption of independence among all the 30 markers is reasonable for the five populations of Qatar.

3.3.4.4 Pairwise F_{ST} values

The pairwise F_{ST} values were also calculated for five populations using INDELs data. The comparison of pairwise F_{ST} values using 30 INDELs among 5 tested populations revealed that Qatari and Yemeni had the lowest pairwise F_{ST} value (0.00052). Pakistani showed the highest pairwise F_{ST} values (0.0108 to 0.03112) in combinations with other populations (Table 3.10).

3.3.4.5 Forensic parameters

The forensic suitability of the DIPplex loci was also evaluated for all the five population groups in Qatar (Table 3.4-3.8). The Combined Power of Discrimination (CPD) for the 30 INDEL loci was 0.9999999 for all of five populations. The Combined Probability of Match (CPM) was calculated in the range of 1.05×10^{-12} (Qatari) to 9.95×10^{-13} (Tunisian). The Combined Probability of Exclusion (CPE) was found to be in the range of 0.9963 (Sudanese and Yemeni) to 0.9974 (Pakistani).

Table.3.4. Table summarizing the forensically relevant parameters of 30 DIPplex INDELs for **Qatari** population, n = 100 (DIP, Deletion Insertion Alleles frequencies, PIC, Polymorphic Information Content; PD, Power of Discrimination; PM, Probability of Match; PE, Power of Exclusion; TPI, Typical Paternity Index; Ho, Observed Heterozygosity; He, Expected Heterozygosity, HWE, *P*-value for Hardy–Weinberg Equilibrium).

QATARI	DIP(-)	DIP(+)	PIC	PD	PM	PE	TPI	Ho	He	HWE
HLD77	0.65	0.35	0.35	0.58	0.42	0.19	1.00	0.50	0.46	0.38
HLD45	0.40	0.60	0.36	0.63	0.37	0.14	0.89	0.44	0.48	0.41
HLD131	0.42	0.59	0.37	0.63	0.37	0.15	0.91	0.45	0.49	0.54
HLD70	0.32	0.69	0.34	0.59	0.41	0.11	0.82	0.39	0.43	0.36
HLD6	0.59	0.42	0.37	0.62	0.38	0.16	0.94	0.47	0.49	0.84
HLD111	0.47	0.53	0.37	0.64	0.36	0.15	0.93	0.46	0.50	0.43
HLD58	0.54	0.46	0.37	0.65	0.35	0.14	0.89	0.44	0.50	0.32
HLD56	0.22	0.78	0.28	0.51	0.49	0.06	0.71	0.30	0.34	0.24
HLD118	0.67	0.33	0.34	0.57	0.43	0.19	1.00	0.50	0.44	0.26
HLD92	0.43	0.57	0.37	0.58	0.42	0.25	1.14	0.56	0.49	0.22
HLD93	0.53	0.47	0.37	0.63	0.37	0.17	0.96	0.48	0.50	0.69
HLD99	0.45	0.56	0.37	0.59	0.41	0.24	1.11	0.55	0.50	0.32
HLD88	0.37	0.63	0.36	0.62	0.38	0.14	0.89	0.44	0.47	0.67
HLD101	0.45	0.56	0.37	0.63	0.37	0.16	0.94	0.47	0.50	0.68
HLD67	0.47	0.54	0.37	0.66	0.34	0.12	0.85	0.41	0.50	0.07
HLD83	0.52	0.49	0.37	0.57	0.43	0.28	1.22	0.59	0.50	0.11
HLD114	0.48	0.53	0.37	0.60	0.41	0.24	1.11	0.55	0.50	0.42
HLD48	0.55	0.45	0.37	0.66	0.34	0.10	0.81	0.38	0.50	0.03
HLD124	0.56	0.45	0.37	0.53	0.47	0.33	1.35	0.63	0.50	0.01
HLD122	0.61	0.40	0.36	0.57	0.43	0.24	1.11	0.55	0.48	0.21
HLD125	0.46	0.55	0.37	0.58	0.42	0.26	1.16	0.57	0.50	0.16
HLD64	0.36	0.65	0.35	0.60	0.40	0.15	0.91	0.45	0.46	0.83
HLD81	0.64	0.37	0.36	0.53	0.47	0.28	1.22	0.59	0.47	0.01
HLD136	0.25	0.75	0.30	0.54	0.46	0.10	0.81	0.38	0.38	1.00
HLD133	0.61	0.40	0.36	0.60	0.40	0.20	1.02	0.51	0.48	0.68
HLD97	0.62	0.38	0.36	0.61	0.39	0.17	0.96	0.48	0.47	1.00
HLD40	0.64	0.36	0.35	0.60	0.40	0.15	0.93	0.46	0.46	1.00
HLD128	0.57	0.44	0.37	0.62	0.38	0.18	0.98	0.49	0.49	1.00
HLD39	0.49	0.52	0.37	0.63	0.37	0.18	0.98	0.49	0.50	0.84
HLD84	0.29	0.71	0.33	0.58	0.42	0.07	0.74	0.32	0.41	0.03

Table.3.5. Table summarizing the forensically relevant parameters of 30 DIPplex INDELs for **Pakistani** population, n = 100 (DIP, Deletion Insertion Alleles frequencies, PIC, Polymorphic Information Content; PD, Power of Discrimination; PM, Probability of Match; PE, Power of Exclusion; TPI, Typical Paternity Index; Ho, Observed Heterozygosity; He, Expected Heterozygosity, HWE, *P*-value for Hardy–Weinberg Equilibrium).

PAKISTANI	DIP(-)	DIP(+)	PIC	PD	PM	PE	TPI	Ho	He	HWE
HLD77	0.62	0.39	0.36	0.60	0.40	0.18	0.98	0.49	0.48	0.83
HLD45	0.31	0.69	0.34	0.59	0.41	0.10	0.81	0.38	0.43	0.25
HLD131	0.47	0.53	0.37	0.66	0.34	0.10	0.81	0.38	0.50	0.02
HLD70	0.36	0.65	0.35	0.62	0.38	0.11	0.82	0.39	0.46	0.13
HLD6	0.52	0.49	0.37	0.66	0.34	0.12	0.85	0.41	0.50	0.07
HLD111	0.53	0.48	0.37	0.63	0.37	0.18	0.98	0.49	0.50	0.84
HLD58	0.59	0.42	0.37	0.62	0.38	0.18	0.98	0.49	0.49	1.00
HLD56	0.35	0.65	0.35	0.62	0.38	0.06	0.71	0.30	0.46	0.01
HLD118	0.66	0.34	0.35	0.60	0.40	0.14	0.89	0.44	0.45	0.83
HLD92	0.46	0.55	0.37	0.65	0.35	0.12	0.85	0.41	0.50	0.11
HLD93	0.61	0.39	0.36	0.62	0.38	0.15	0.93	0.46	0.48	0.83
HLD99	0.39	0.62	0.36	0.63	0.37	0.12	0.85	0.41	0.48	0.21
HLD88	0.47	0.53	0.37	0.66	0.34	0.10	0.81	0.38	0.50	0.02
HLD101	0.44	0.57	0.37	0.61	0.39	0.20	1.02	0.51	0.49	0.84
HLD67	0.49	0.51	0.37	0.61	0.39	0.21	1.04	0.52	0.50	0.84
HLD83	0.70	0.30	0.33	0.58	0.42	0.13	0.86	0.42	0.42	1.00
HLD114	0.66	0.35	0.35	0.59	0.41	0.16	0.94	0.47	0.45	0.83
HLD48	0.57	0.44	0.37	0.64	0.36	0.15	0.91	0.45	0.49	0.42
HLD124	0.54	0.47	0.37	0.38	0.62	0.54	2.17	0.77	0.50	0.01
HLD122	0.51	0.49	0.37	0.60	0.40	0.22	1.09	0.54	0.50	0.55
HLD125	0.38	0.63	0.36	0.52	0.48	0.30	1.28	0.61	0.47	0.01
HLD64	0.25	0.75	0.30	0.54	0.46	0.07	0.74	0.32	0.38	0.18
HLD81	0.55	0.46	0.37	0.45	0.55	0.44	1.72	0.71	0.50	0.01
HLD136	0.43	0.58	0.37	0.61	0.39	0.20	1.02	0.51	0.49	0.84
HLD133	0.64	0.36	0.35	0.62	0.38	0.13	0.86	0.42	0.46	0.39
HLD97	0.54	0.46	0.37	0.64	0.36	0.15	0.93	0.46	0.50	0.55
HLD40	0.68	0.32	0.34	0.60	0.40	0.10	0.81	0.38	0.44	0.25
HLD128	0.55	0.46	0.37	0.60	0.40	0.22	1.06	0.53	0.50	0.55
HLD39	0.45	0.56	0.37	0.65	0.35	0.12	0.85	0.41	0.50	0.10
HLD84	0.44	0.57	0.37	0.64	0.36	0.15	0.91	0.45	0.49	0.42

Table.3.6. Table summarizing the forensically relevant parameters of 30 DIPplex INDELs for Sudanese population, n = 100 (DIP, Deletion Insertion Alleles frequencies, PIC, Polymorphic Information Content; PD, Power of Discrimination; PM, Probability of Match; PE, Power of Exclusion; TPI, Typical Paternity Index; Ho, Observed Heterozygosity; He, Expected Heterozygosity, HWE, *P*-value for Hardy–Weinberg Equilibrium).

SUDANESE	DIP(-)	DIP(+)	PIC	PD	PM	PE	TPI	Ho	He	HWE
HLD77	0.73	0.27	0.32	0.56	0.44	0.07	0.74	0.32	0.40	0.07
HLD45	0.46	0.55	0.37	0.58	0.42	0.26	1.16	0.57	0.50	0.16
HLD131	0.39	0.61	0.36	0.59	0.41	0.21	1.04	0.52	0.48	0.41
HLD70	0.29	0.72	0.32	0.57	0.43	0.05	0.68	0.27	0.41	0.01
HLD6	0.70	0.30	0.33	0.57	0.43	0.14	0.89	0.44	0.42	0.81
HLD111	0.50	0.51	0.37	0.64	0.36	0.16	0.94	0.47	0.50	0.55
HLD58	0.72	0.29	0.32	0.57	0.43	0.11	0.82	0.39	0.41	0.63
HLD56	0.31	0.69	0.34	0.59	0.41	0.11	0.83	0.40	0.43	0.49
HLD118	0.60	0.40	0.36	0.58	0.42	0.22	1.09	0.54	0.48	0.29
HLD92	0.53	0.48	0.37	0.62	0.38	0.20	1.02	0.51	0.50	1.00
HLD93	0.59	0.41	0.37	0.54	0.46	0.29	1.25	0.60	0.49	0.02
HLD99	0.23	0.78	0.29	0.51	0.49	0.07	0.72	0.31	0.35	0.26
HLD88	0.29	0.71	0.33	0.58	0.42	0.09	0.78	0.36	0.41	0.22
HLD101	0.38	0.62	0.36	0.63	0.37	0.13	0.86	0.42	0.47	0.29
HLD67	0.48	0.53	0.37	0.65	0.36	0.15	0.91	0.45	0.50	0.32
HLD83	0.55	0.45	0.37	0.62	0.38	0.19	1.00	0.50	0.50	1.00
HLD114	0.40	0.61	0.36	0.61	0.39	0.18	0.98	0.49	0.48	1.00
HLD48	0.34	0.67	0.35	0.58	0.42	0.18	0.98	0.49	0.45	0.38
HLD124	0.68	0.33	0.34	0.56	0.44	0.20	1.02	0.51	0.44	0.17
HLD122	0.67	0.34	0.35	0.58	0.42	0.16	0.94	0.47	0.45	0.66
HLD125	0.53	0.47	0.37	0.46	0.54	0.43	1.67	0.70	0.50	0.01
HLD64	0.25	0.75	0.30	0.54	0.46	0.06	0.69	0.28	0.38	0.02
HLD81	0.63	0.38	0.36	0.58	0.42	0.22	1.06	0.53	0.47	0.28
HLD136	0.29	0.71	0.33	0.57	0.43	0.13	0.86	0.42	0.41	1.00
HLD133	0.52	0.49	0.37	0.63	0.37	0.18	0.98	0.49	0.50	0.84
HLD97	0.57	0.43	0.37	0.57	0.43	0.27	1.19	0.58	0.49	0.10
HLD40	0.65	0.35	0.35	0.61	0.39	0.13	0.86	0.42	0.46	0.51
HLD128	0.41	0.60	0.37	0.58	0.42	0.24	1.11	0.55	0.48	0.21
HLD39	0.54	0.47	0.37	0.65	0.35	0.13	0.88	0.43	0.50	0.17
HLD84	0.18	0.83	0.25	0.45	0.56	0.04	0.67	0.25	0.29	0.17

Table.3.7. Table summarizing the forensically relevant parameters of 30 DIPplex INDELs for Yemeni population, n = 100 (DIP, Deletion Insertion Alleles frequencies, PIC, Polymorphic Information Content; PD, Power of Discrimination; PM, Probability of Match; PE, Power of Exclusion; TPI, Typical Paternity Index; Ho, Observed Heterozygosity; He, Expected Heterozygosity, HWE, *P*-value for Hardy–Weinberg Equilibrium).

YEMENI	DIP(-)	DIP(+)	PIC	PD	PM	PE	TPI	Ho	He	HWE
HLD77	0.71	0.29	0.33	0.57	0.43	0.12	0.85	0.41	0.42	1.00
HLD45	0.46	0.54	0.37	0.66	0.34	0.09	0.78	0.36	0.50	0.01
HLD131	0.44	0.56	0.37	0.63	0.37	0.17	0.96	0.48	0.50	0.84
HLD70	0.43	0.57	0.37	0.62	0.38	0.17	0.96	0.48	0.49	0.84
HLD6	0.51	0.49	0.37	0.66	0.34	0.11	0.82	0.39	0.50	0.03
HLD111	0.49	0.51	0.37	0.63	0.37	0.18	0.98	0.49	0.50	0.84
HLD58	0.63	0.37	0.36	0.58	0.42	0.21	1.04	0.52	0.47	0.29
HLD56	0.19	0.81	0.26	0.47	0.53	0.05	0.68	0.27	0.32	0.20
HLD118	0.69	0.31	0.34	0.56	0.44	0.17	0.96	0.48	0.43	0.35
HLD92	0.42	0.58	0.37	0.57	0.43	0.26	1.16	0.57	0.49	0.10
HLD93	0.51	0.49	0.37	0.58	0.42	0.26	1.16	0.57	0.50	0.23
HLD99	0.45	0.55	0.37	0.63	0.37	0.16	0.94	0.47	0.50	0.69
HLD88	0.36	0.64	0.35	0.63	0.37	0.07	0.74	0.32	0.46	0.01
HLD101	0.51	0.49	0.37	0.66	0.34	0.13	0.86	0.42	0.50	0.11
HLD67	0.45	0.55	0.37	0.65	0.35	0.13	0.86	0.42	0.50	0.16
HLD83	0.54	0.46	0.37	0.60	0.40	0.22	1.09	0.54	0.50	0.43
HLD114	0.57	0.43	0.37	0.62	0.38	0.17	0.96	0.48	0.49	0.84
HLD48	0.57	0.43	0.37	0.62	0.38	0.19	1.00	0.50	0.49	1.00
HLD124	0.47	0.53	0.37	0.57	0.43	0.28	1.22	0.59	0.50	0.11
HLD122	0.60	0.40	0.37	0.64	0.36	0.12	0.85	0.41	0.48	0.15
HLD125	0.51	0.49	0.37	0.49	0.51	0.40	1.56	0.68	0.50	0.01
HLD64	0.36	0.64	0.35	0.62	0.38	0.11	0.83	0.40	0.46	0.20
HLD81	0.63	0.37	0.36	0.61	0.39	0.15	0.93	0.46	0.47	1.00
HLD136	0.28	0.72	0.32	0.57	0.43	0.08	0.76	0.34	0.41	0.14
HLD133	0.61	0.39	0.36	0.54	0.46	0.29	1.25	0.60	0.48	0.01
HLD97	0.61	0.39	0.36	0.62	0.38	0.15	0.93	0.46	0.48	0.83
HLD40	0.56	0.45	0.37	0.65	0.35	0.12	0.85	0.41	0.50	0.10
HLD128	0.56	0.44	0.37	0.58	0.42	0.25	1.14	0.56	0.50	0.22
HLD39	0.57	0.43	0.37	0.65	0.35	0.11	0.82	0.39	0.49	0.04
HLD84	0.27	0.73	0.32	0.56	0.44	0.07	0.74	0.32	0.40	0.07

Table.3.8. Table summarizing the forensically relevant parameters of 30 DIPplex INDELs for Tunisian population, n = 100 (DIP, Deletion Insertion Alleles frequencies, PIC, Polymorphic Information Content; PD, Power of Discrimination; PM, Probability of Match; PE, Power of Exclusion; TPI, Typical Paternity Index; Ho, Observed Heterozygosity; He, Expected Heterozygosity, HWE, *P*-value for Hardy–Weinberg Equilibrium).

TUNISIAN	DIP(-)	DIP(+)	PIC	PD	PM	PE	TPI	Ho	He	HWE
HLD77	0.61	0.39	0.36	0.63	0.37	0.13	0.86	0.42	0.48	0.29
HLD45	0.54	0.46	0.37	0.63	0.37	0.17	0.96	0.48	0.50	0.84
HLD131	0.41	0.59	0.37	0.62	0.38	0.16	0.94	0.47	0.48	0.84
HLD70	0.28	0.72	0.32	0.57	0.43	0.10	0.81	0.38	0.41	0.62
HLD6	0.61	0.39	0.36	0.62	0.38	0.16	0.94	0.47	0.48	0.84
HLD111	0.48	0.52	0.37	0.62	0.38	0.20	1.02	0.51	0.50	1.00
HLD58	0.65	0.35	0.35	0.60	0.40	0.14	0.89	0.44	0.46	0.83
HLD56	0.30	0.70	0.33	0.58	0.42	0.13	0.86	0.42	0.42	1.00
HLD118	0.67	0.33	0.34	0.60	0.40	0.13	0.86	0.42	0.44	0.65
HLD92	0.50	0.50	0.38	0.63	0.38	0.19	1.00	0.50	0.50	1.00
HLD93	0.50	0.50	0.38	0.63	0.38	0.19	1.00	0.50	0.50	1.00
HLD99	0.35	0.65	0.35	0.60	0.40	0.16	0.94	0.47	0.46	1.00
HLD88	0.34	0.66	0.35	0.61	0.39	0.10	0.79	0.37	0.45	0.11
HLD101	0.45	0.55	0.37	0.57	0.43	0.27	1.19	0.58	0.50	0.11
HLD67	0.37	0.63	0.36	0.59	0.41	0.18	0.98	0.49	0.47	0.67
HLD83	0.55	0.45	0.37	0.64	0.36	0.15	0.93	0.46	0.50	0.54
HLD114	0.51	0.49	0.37	0.65	0.35	0.14	0.89	0.44	0.50	0.24
HLD48	0.45	0.55	0.37	0.63	0.37	0.18	0.98	0.49	0.50	1.00
HLD124	0.55	0.45	0.37	0.55	0.45	0.30	1.28	0.61	0.50	0.03
HLD122	0.59	0.41	0.37	0.65	0.35	0.10	0.79	0.37	0.49	0.02
HLD125	0.44	0.56	0.37	0.45	0.55	0.44	1.72	0.71	0.49	0.01
HLD64	0.33	0.67	0.34	0.59	0.41	0.13	0.88	0.43	0.44	0.82
HLD81	0.69	0.31	0.34	0.52	0.48	0.25	1.14	0.56	0.43	0.01
HLD136	0.29	0.71	0.33	0.58	0.42	0.09	0.78	0.36	0.41	0.23
HLD133	0.52	0.48	0.37	0.63	0.37	0.18	0.98	0.49	0.50	0.84
HLD97	0.58	0.42	0.37	0.62	0.38	0.18	0.98	0.49	0.49	1.00
HLD40	0.65	0.35	0.35	0.62	0.38	0.07	0.72	0.31	0.46	0.01
HLD128	0.51	0.49	0.37	0.63	0.37	0.17	0.96	0.48	0.50	0.69
HLD39	0.64	0.36	0.35	0.60	0.40	0.17	0.96	0.48	0.46	0.83
HLD84	0.30	0.70	0.33	0.59	0.41	0.07	0.72	0.31	0.43	0.01

Table 3.9 Table showing the significant linkage disequilibrium at $P < 0.05$ in five populations of Qatar.

Marker code	LOCI	Qatari	Pakistani	Sudanese	Tunisian	Yemeni
0	**HLD77**	1	4	2	1	4
1	**HLD45**	2	1	3	1	1
2	**HLD131**	2	4	1	1	1
3	**HLD70**	2	5	3	5	3
4	**HLD6**	0	0	1	1	4
5	**HLD111**	6	1	4	5	1
6	**HLD58**	2	0	0	0	4
7	**HLD56**	3	3	1	1	1
8	**HLD118**	2	4	3	2	3
9	**HLD92**	4	3	2	0	2
10	**HLD93**	4	2	4	2	1
11	**HLD99**	5	2	0	1	3
12	**HLD88**	3	3	4	0	1
13	**HLD101**	2	2	1	1	4
14	**HLD67**	2	1	1	0	3
15	**HLD83**	0	1	3	4	2
16	**HLD114**	0	3	3	5	2
17	**HLD48**	1	2	0	0	3
18	**HLD124**	3	2	4	4	2
19	**HLD122**	3	5	2	2	2
20	**HLD125**	4	0	3	0	3
21	**HLD64**	4	4	3	1	1
22	**HLD81**	1	3	0	0	2
23	**HLD136**	1	0	0	1	6
24	**HLD133**	1	0	1	1	1
25	**HLD97**	1	2	2	0	3
26	**HLD40**	1	3	2	2	3
27	**HLD128**	1	2	1	0	2
28	**HLD39**	1	2	4	3	1
29	**HLD84**	2	4	2	2	3
Total linked loci		64	68	60	46	72
		32pairs	34 pairs	30 pairs	23 pairs	36pairs

Table 3.10 Table showing the pairwise F_{ST} values between 5 tested population based on 30 INDELs (*result not significant).

	Yemeni	Tunisian	Sudanese	Qatari	Pakistani
Yemeni	0				
Tunisian	0.00553	0			
Sudanese	0.02201	0.00892	0		
Qatari	0.00052*	0.00359	0.01511	0	
Pakistani	0.01356	0.01673	0.03112	0.0108	0

3.3.5 Statistical Analysis of STRs Data

3.3.5.1 HWE

While calculating p- value for Hardy-Weinberg equilibrium, out of total 15 loci, 3 (Qataris), 2 (Pakistanis) and 4 (Sudanese) loci showed departure from p-value. But, after applying Bonferroni's correction at $P < 0.0033$ (0.05/15), there was no deviation from Hardy-Weinberg equilibrium. The expected heterozygosities ranged from 0.610 (Yemeni) to 0.890 (Sudanese and Tunisian) with a mean value of 0.750, while the observed heterozygosities ranged from 0.590 (Yemeni) to 0.900 (Sudanese) with a mean value of 0.745. The values for the polymorphic information content ranged between 0.620 (Qatari and Pakistani) and 0.890 (Sudanese). Tables 3.11-3.15 describe HWE and other statistical parameters for five populations. Figure 3.14 shows the values of heterozygosities (observed and expected) for FGA locus.

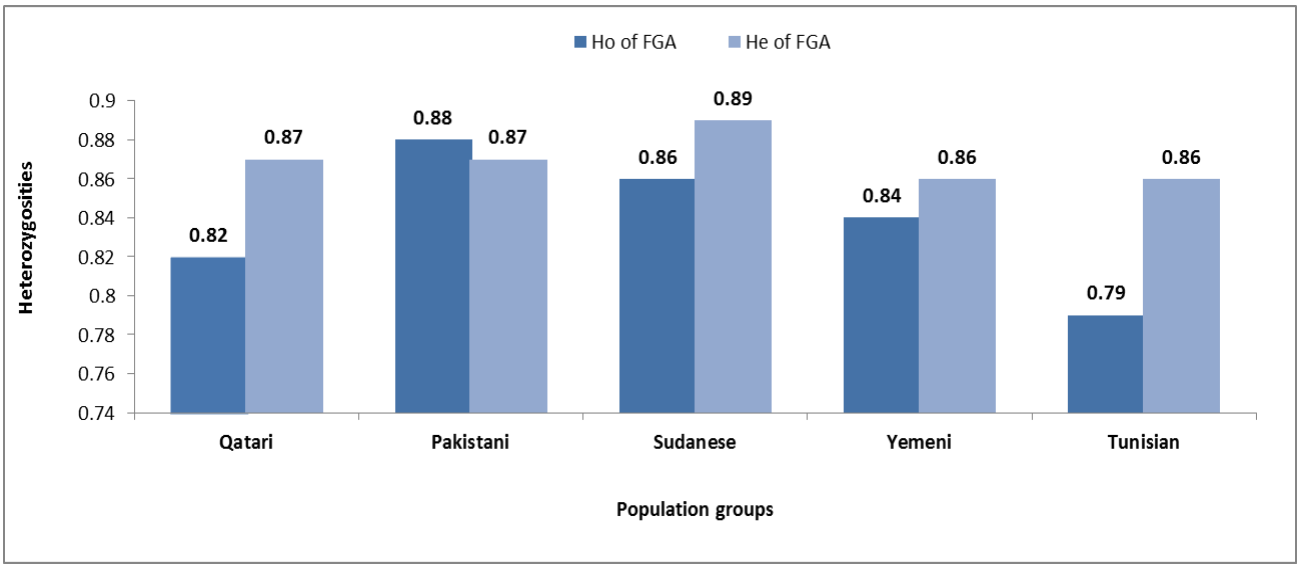

Figure 3.14 Bar chart showing the values of observed and expected heterozygosities for FGA locus.

3.3.5.2 Forensic parameters

The forensic suitability of the 15 STR loci was also evaluated for all the five population groups in Qatar (Table 3.18). The Combined Power of Discrimination (CPD) was 0.9999999 for all of five populations. The Combined Probability of Match (CPM) was calculated in the range of

1.90 x 10^{-20} (Qatari) to 9.45 x 10^{-21} (Tunisian). The Combined Probability of Exclusion (CPE) was found to be in the range of 0.99999717 (Sudanese) to 0.99999903 (Pakistani).

3.3.5.3 Pairwise F_{ST} values

The pairwise F_{ST} values calculated for STRs showed the same pattern as INDELs. Qatari and Yemeni showed a close relatedness based on the pairwise F_{ST} value (0.00155). The Pakistani population again showed the highest pairwise F_{ST} values (0.00745 to 0.01199). Table 3.16 illustrates the pairwise F_{ST} values for 5 tested populations using STR data.

Table.3.11. Table showing the forensically relevant parameters of 15 STRs for Qatari population, n = 100 (PM, Probability of Match; PD, Power of Discrimination; PIC, Polymorphic Information Content; PE, Power of Exclusion; TPI, Typical Paternity Index; Ho, Observed Heterozygosity; He, Expected Heterozygosity, HWE, P-value for Hardy–Weinberg Equilibrium).

QATARI	PM	PD	PIC	PE	TPI	Ho	He	HWE
D8S1179	0.14	0.86	0.68	0.58	2.38	0.89	0.85	0.26
D21S11	0.06	0.94	0.82	0.78	4.55	0.82	0.83	0.09
D7S820	0.06	0.94	0.81	0.64	2.78	0.76	0.77	0.88
CSFIPO	0.14	0.86	0.68	0.58	2.38	0.79	0.73	0.25
D3S1358	0.10	0.90	0.71	0.27	1.19	0.58	0.76	0.01
TH01	0.12	0.88	0.71	0.56	2.27	0.78	0.75	0.23
D13S317	0.09	0.91	0.73	0.46	1.79	0.72	0.76	0.51
D16S539	0.09	0.91	0.72	0.48	1.85	0.73	0.76	0.49
D2S1338	0.04	0.96	0.86	0.69	3.33	0.85	0.88	0.01
D19S433	0.05	0.95	0.83	0.66	2.94	0.83	0.85	0.81
VWA	0.07	0.93	0.78	0.66	2.94	0.83	0.81	0.93
TPOX	0.16	0.84	0.62	0.34	1.39	0.64	0.67	0.58
D18S51	0.04	0.96	0.86	0.62	2.63	0.81	0.87	0.01
D5S818	0.09	0.91	0.72	0.36	1.43	0.65	0.77	0.09
FGA	0.04	0.96	0.85	0.64	2.78	0.82	0.87	0.66

Table.3.12. Table showing the forensically relevant parameters of 15 STRs for Pakistani population, n = 100 (PM, Probability of Match; PD, Power of Discrimination; PIC, Polymorphic Information Content; PE, Power of Exclusion; TPI, Typical Paternity Index; Ho, Observed Heterozygosity; He, Expected Heterozygosity, HWE, P-value for Hardy–Weinberg Equilibrium).

PAKISTANI	PM	PD	PIC	PE	TPI	Ho	He	HWE
D8S1179	0.14	0.86	0.69	0.54	2.17	0.89	0.85	0.64
D21S11	0.05	0.95	0.82	0.78	4.55	0.89	0.87	0.44
D7S820	0.04	0.96	0.85	0.78	4.55	0.79	0.81	0.17
CSFIPO	0.14	0.86	0.69	0.54	2.17	0.77	0.74	0.04
D3S1358	0.11	0.89	0.71	0.51	2.00	0.75	0.75	0.96
TH01	0.08	0.92	0.75	0.54	2.17	0.77	0.79	0.61
D13S317	0.07	0.93	0.77	0.44	1.72	0.71	0.80	0.37
D16S539	0.10	0.90	0.72	0.41	1.61	0.69	0.76	0.04
D2S1338	0.04	0.96	0.86	0.64	2.78	0.82	0.88	0.22
D19S433	0.06	0.94	0.80	0.62	2.63	0.81	0.83	0.35
VWA	0.06	0.94	0.80	0.62	2.63	0.81	0.83	0.64
TPOX	0.15	0.85	0.62	0.38	1.52	0.67	0.68	0.63
D18S51	0.04	0.96	0.85	0.71	3.57	0.86	0.87	0.14
D5S818	0.12	0.88	0.69	0.44	1.72	0.71	0.74	0.83
FGA	0.04	0.96	0.85	0.75	4.17	0.88	0.87	0.86

Table.3.13. Table showing the forensically relevant parameters of 15 STRs for Sudanese population, n = 100 (PM, Probability of Match; PD, Power of Discrimination; PIC, Polymorphic Information Content; PE, Power of Exclusion; TPI, Typical Paternity Index; Ho, Observed Heterozygosity; He, Expected Heterozygosity, HWE, P-value for Hardy–Weinberg Equilibrium).

SUDANESE	PM	PD	PIC	PE	TPI	Ho	He	HWE
D8S1179	0.14	0.86	0.68	0.51	2.00	0.80	0.82	0.41
D21S11	0.06	0.94	0.80	0.60	2.50	0.74	0.82	0.01
D7S820	0.06	0.94	0.80	0.49	1.92	0.80	0.78	0.25
CSFIPO	0.14	0.86	0.68	0.51	2.00	0.75	0.73	0.11
D3S1358	0.11	0.89	0.70	0.51	2.00	0.75	0.75	0.41
TH01	0.13	0.87	0.68	0.53	2.08	0.76	0.72	0.73
D13S317	0.08	0.92	0.74	0.48	1.85	0.73	0.78	0.52
D16S539	0.08	0.92	0.76	0.46	1.79	0.72	0.80	0.03
D2S1338	0.04	0.96	0.86	0.80	5.00	0.90	0.88	0.14
D19S433	0.06	0.94	0.81	0.56	2.27	0.78	0.83	0.02
VWA	0.07	0.94	0.79	0.62	2.63	0.81	0.82	0.10
TPOX	0.16	0.84	0.64	0.41	1.61	0.69	0.70	0.20
D18S51	0.03	0.97	0.89	0.71	3.57	0.86	0.90	0.40
D5S818	0.08	0.92	0.75	0.46	1.79	0.72	0.78	0.43
FGA	0.04	0.96	0.87	0.71	3.57	0.86	0.89	0.01

Table.3.14. Table showing the forensically relevant parameters of 15 STRs for Yemeni population, n = 100 (PM, Probability of Match; PD, Power of Discrimination; PIC, Polymorphic Information Content; PE, Power of Exclusion; TPI, Typical Paternity Index; Ho, Observed Heterozygosity; He, Expected Heterozygosity, HWE, P-value for Hardy–Weinberg Equilibrium).

YEMENI	PM	PD	PIC	PE	TPI	Ho	He	HWE
D8S1179	0.15	0.85	0.66	0.56	2.27	0.77	0.84	0.13
D21S11	0.05	0.95	0.82	0.54	2.17	0.82	0.85	0.62
D7S820	0.05	0.95	0.83	0.64	2.78	0.81	0.75	0.45
CSFIPO	0.15	0.85	0.66	0.56	2.27	0.78	0.72	0.32
D3S1358	0.10	0.90	0.73	0.51	2.00	0.75	0.77	0.29
TH01	0.10	0.90	0.74	0.56	2.27	0.78	0.78	0.26
D13S317	0.08	0.92	0.75	0.60	2.50	0.80	0.78	0.37
D16S539	0.11	0.89	0.71	0.40	1.56	0.68	0.74	0.11
D2S1338	0.04	0.96	0.87	0.69	3.33	0.85	0.88	0.20
D19S433	0.05	0.95	0.84	0.66	2.94	0.83	0.86	0.24
VWA	0.09	0.91	0.75	0.60	2.50	0.80	0.79	0.14
TPOX	0.20	0.80	0.56	0.28	1.22	0.59	0.61	0.32
D18S51	0.04	0.96	0.85	0.58	2.38	0.79	0.87	0.20
D5S818	0.12	0.88	0.71	0.62	2.63	0.81	0.75	0.23
FGA	0.04	0.96	0.84	0.68	3.13	0.84	0.86	0.06

Table.3.15. Table showing the forensically relevant parameters of 15 STRs for Tunisian population, n = 100 (PM, Probability of Match; PD, Power of Discrimination; PIC, Polymorphic Information Content; PE, Power of Exclusion; TPI, Typical Paternity Index; Ho, Observed Heterozygosity; He, Expected Heterozygosity, HWE, P-value for Hardy–Weinberg Equilibrium).

TUNISIAN	PM	PD	PIC	PE	TPI	Ho	He	HWE
D8S1179	0.12	0.88	0.68	0.41	1.61	0.83	0.82	0.95
D21S11	0.06	0.94	0.79	0.66	2.94	0.88	0.83	0.83
D7S820	0.06	0.94	0.81	0.75	4.17	0.81	0.79	0.77
CSFIPO	0.12	0.88	0.68	0.41	1.61	0.69	0.73	0.46
D3S1358	0.12	0.88	0.71	0.62	2.63	0.81	0.76	0.69
TH01	0.07	0.93	0.77	0.41	1.61	0.69	0.80	0.12
D13S317	0.10	0.90	0.72	0.40	1.56	0.68	0.76	0.18
D16S539	0.08	0.92	0.74	0.51	2.00	0.75	0.78	0.45
D2S1338	0.04	0.96	0.84	0.73	3.85	0.87	0.85	0.78
D19S433	0.07	0.93	0.77	0.56	2.27	0.78	0.81	0.26
VWA	0.07	0.93	0.78	0.60	2.50	0.80	0.82	0.62
TPOX	0.13	0.87	0.68	0.48	1.85	0.73	0.73	0.06
D18S51	0.03	0.97	0.88	0.71	3.57	0.86	0.89	0.55
D5S818	0.12	0.88	0.69	0.53	2.08	0.76	0.73	0.85
FGA	0.04	0.96	0.84	0.58	2.38	0.79	0.86	0.16

Table 3.16 Table showing the pairwise F_{ST} values between 5 tested population based on 15 STRs (*result not significant).

	Yemeni	Tunisian	Sudanese	Qatari	Pakistani
Yemeni	0				
Tunisian	0.00662	0			
Sudanese	0.01087	0.00451	0		
Qatari	0.00155*	0.00371	0.00851	0	
Pakistani	0.00850	0.00810	0.01199	0.00745	0

3.4. Discussion

In this study, 500 samples were profiled using the Investigator® DIPplex kit (Qiagen™). The purpose was to evaluate the forensic efficiency of autosomal INDELs with the possibility of using them for forensic casework. In order to achieve this goal, different forensic/population parameters were calculated by using the genetic data generated from 500 DNA samples from volunteers belonging to five different nationalities (Qatari, Pakistani, Sudanese, Yemeni and Tunisian).

The PCR optimisation of Investigator® DIPplex kit was done in terms of its sensitivity and reaction volume. The allele balance and peak quality was better for 0.5 ng of input DNA template which was also confirmed from the results of previous studies done by using Investigator® DIPplex kit (Neuvonen *et al.* 2011; Martin *et al.* 2012; Akhteruzzaman *et al.* 2013; Nunotani *et al.* 2015). The validation of reaction volume was performed for one half, one quarter and one fifth of the recommended PCR reaction mix of 25 µl. The results indicated that one half and one quarter of the recommended volume could be used to generate quality profiles. A modification in the protocol of Investigator® DIPplex kit during capillary electrophoresis was made by using different internal size standard of LIZ-500 (Applied Biosystems™) instead of BTO-550 (Qiagen™). The reason for not using BTO-550 was that the sizing was not accurate for samples in different batches. On the other hand, when LIZ-500 was used instead of BTO-550, the accurate sizing of fragments was achieved. After optimising with LIZ-500, then all PCR products of Investigator® DIPplex kit were processed for capillary electrophoresis using Applied Biosystems™ size standard.

The genetic variation of 30 INDEL markers in Investigator® DIPplex kit for the five populations of Qatar are summarized in Tables 3.4-3.9. All 30 INDEL markers contained within Investigator® DIPplex kit were in Hardy-Weinberg equilibrium. The average observed heterozygousity of 30 INDELs was 0.51 in the studied populations which indicates their highly polymorphic nature. Similar results were obtained from other world populations using INDEL markers (Wei *et al.* 2013; Ferreira *et al.* 2015).

The linkage disequilibrium (LD) analysis was performed to assess that the INDEL markers are free of association. LD test did not give any significant *p*-value after Bonferroni's correction. Therefore, INDEL markers can be considered as independent markers to use for statistical calculations for forensic purposes (Da Silva *et al.* 2013; Wei *et al.* 2013; Pepinski *et al.* 2013; Ferreira *et al* 2015). For Investigator® DIPplex kit, no linkage had been detected from any published population data, but two of STR loci (VWA and D12S391) were identified which are only 6.36 Mb apart on the short arm of chromosome 12 (O'Connor *et al.* 2011). In this type of case, there might be an issue in case of related individuals such as in kinship analysis as product rule cannot be applied to get useful statistical calculation (O'Connor and Tillmar 2012).

The forensic efficiency of 30 INDELs in Investigator® DIPplex kit for five populations based in Qatar was calculated in terms of combined discrimination power (CDP), combined probability match (CPM) and combined power of exclusion (CPE). These values were then compared with those of published populations (Table 3.17). The Combined Power of Discrimination (CPD) for the 30 INDEL loci was 0.9999999 for all of five populations which is a satisfactory value for forensic purpose. The Combined Probability of Match (CPM) was calculated in the range of 10^{-12} (Qatari) to 10^{-13} (Tunisian) and the Combined Probability of Exclusion (CPE) was found to be in the range of 0.9963 (Sudanese and Yemeni) to 0.9974 (Pakistani) which are comparable to other world populations (Table 3.17).

A comparison of the CPM and CPE values for the 30 INDEL markers and 15 STR markers for the same 500 samples of all the five studied population groups in Qatar was also done (Table 3.18) which showed that both values for INDELs are much lower than those of STR loci, which was expected due to biallelic nature of INDELs as compared to STRs which have multiple alleles in a single locus.

Table.3.17. Table showing the comparison of DIPplex data between different sub-population groups of Qatar and other world populations.

Multiplex	Populations	CPM	CPE	References
DIPplex	Qatari	1.06×10^{-12}	0.9973	----------
DIPplex	Pakistani	8.90×10^{-13}	0.9974	----------
DIPplex	Sudanese	5.57×10^{-12}	0.9963	----------
DIPplex	Yemeni	7.92×10^{-13}	0.9963	----------
DIPplex	Tunisian	9.95×10^{-13}	0.9964	----------
DIPplex	Chinese	1.28×10^{-12}	0.9968	Wang *et al*. 2015
DIPplex	Brazilian	3.40×10^{-13}	0.9973	Ferreira *et al.* 2015
DIPplex	Japanese	2.67×10^{-11}	0.9964	Nunotani *et al.* 2015
DIPplex	S. Korean	2.84×10^{-11}	0.9888	Seong *et al.* 2013
DIPplex	Poles	7.98×10^{-14}	0.9900	Pepinski *et al.* 2013
DIPplex	Taiwanese	1.22×10^{-11}	0.9884	Pepinski *et al*. 2013
DIPplex	Bangladeshi	2.08×10^{-12}	0.9947	Akhteruzzaman *et al*. 2013
DIPplex	Chinese	9.13×10^{-12}	0.9929	Liang *et al*. 2013
DIPplex	Portugal	5.58×10^{-13}	0.9982	Da Silva *et al.* 2013
DIPplex	Spain	5.72×10^{-13}	0.9985	Martin *et al.* 2013
DIPplex	Basque	3.14×10^{-13}	0.9970	Martin *et al.* 2013
DIPplex	Uruguayan	3.21×10^{-13}	0.9971	Saiz *et al.* 2012
DIPplex	Somali	5.03×10^{-12}	0.9620	Neuonen *et al.* 2011
DIPplex	Finnish	3.54×10^{-13}	0.9961	Neuonen *et al.* 2011
DIPplex	Danes	3.30×10^{-13}	0.9970	Friis *et al.* 2011

Table.3.18. Table showing the comparison of 30 INDELs with 15 autosomal STRs between different population groups of Qatar.

	30 INDELs		15 STRs	
	CPM	**CPE**	**CPM**	**CPE**
Qatari	1.06×10^{-12}	0.9973	1.90×10^{-20}	0.99999722
Pakistani	8.90×10^{-13}	0.9974	7.21×10^{-21}	0.99999903
Sudanese	5.57×10^{-12}	0.9963	6.68×10^{-21}	0.99999717
Yemeni	7.92×10^{-13}	0.9963	3.92×10^{-20}	0.99999741
Tunisian	9.95×10^{-13}	0.9964	9.45×10^{-21}	0.99999734

The disparity of pairwise F_{ST} values from one polymorphism to another among populations may help to establish whether genetic drift is playing role or whether variation is selectively natural (Bowcock *et al.* 1994). The low pairwise F_{ST} value between Qatari and Yemeni showed their close relatedness which is most likely explained by common ancestors, i.e. nomads from Arabian Peninsula (Carol *et al.* 2001). This close genetic relationship between Qatari and Yemeni populations was also supported by a Y-chromosome study which indicated Qatari and Yemeni shared most of Y haplogroups together (Cadenas *et al.* 2008). Pakistani and Sudanese showed the highest pairwise F_{ST} values in any of the combinations with other populations which indicated genetic differentiation among them, due to their distant geographical locations (Monica *et al.* 2005). Another important observation was made while comparing the F_{ST} values of INDELs and STRs that the values of F_{ST} for STRs were lower than those of INDELs. The reason for this difference lies in the polymorphic nature of both systems as STRs are more polymorphic than INDELs; therefore their values were lesser than those of INDELs or vice versa (Excoffier *et al*. 2009).

3.5. Conclusion

In conclusion, it can be said that autosomal INDEL markers can be a useful tool for forensic investigation based on their strong statistical values. They might be very helpful in solving the cases where challenging samples are involved (Carvalho *et al*. 2011; Pereira *et al.* 2012). They can also be used as a complementary tool for kinship analysis with STRs, especially in cases involving mutation (Karlsson *et al.* 2007).

CHAPTER 4
ESTIMATION OF GEOGRAPHIC ORIGIN USING STRs AND INDELs

4.1. OVERVIEW

Estimation of ancestry from an unknown crime scene sample can provide valuable leads when there is no eyewitness or hit in a DNA database. Ancestry analysis depends upon the variation in an individual that can help to trace their origin to a particular geographic area (Rohlfs *et al.* 2012). Most of the research for ancestry analysis has been focused on SNPs and INDELs.

Much research has been done on DNA polymorphism for human identification purpose (Chakraborty *et al.* 1999). Most data on forensic markers relates to short tandem repeat (STR) polymorphisms. Due to their high discriminating power, data is available from many different populations from different parts of the world (Butler 2005). In addition to identification, some attention of forensic researchers has moved towards other forensic applications, such as estimation of ancestry (Shriver *et al.* 1997). For this particular application, STRs are not ideal genetic markers due to their high mutation rate (Phillips *et al.* 2007).

As an alternate, single nucleotide polymorphism (SNPs) are considered a better ancestry tool as compared to STRs due to their low mutation rate, common occurrence in human genome, availability of their allele frequencies for a number of populations and short amplicons that can be useful in typing the degraded samples (Frudakis *et al.* 2003; Fondevila *et al.* 2008; Myles *et al.* 2009; Lundsberg *et al*. 2013). However, due to the involvement of multiple steps and complex interpretation for their typing, SNPs were not adopted as a widespread genetic tool to derive ancestry information in routine forensic laboratories (Phillips *et al.* 2007).

Insertion Deletion polymorphisms (INDELs) have emerged as another forensic tool to be used as biogeographic ancestry tool (Fridakis *et al*. 2003; Phillips *et al*. 2007; Kayser and De Kniff 2011; Pereira *et al.* 2012; Francez *et al.* 2012; Romanini *et al.* 2014). INDELs combine the advantages of both STRs and SNPs, which include low mutation rate, high abundance in human genome

(Mills *et al.* 2006) and different allele frequency distributions among different geographical populations (Nachman *et al.* 2000). INDEL based Ancestry panels have been developed and widely used for different world populations. Some AIM-INDEL panels developed include 48 INDELs in three multiplexes (Santos *et al.* 2010), 48 INDELs in one multiplex (Pereira *et al.* 2012), and 21 INDELs in one multiplex (Zaumsegel *et al.* 2013). Although AIM-INDELs have not the same power as AIM-SNPs, they have the advantage that capillary electrophoresis is used to detect length polymorphism, which is similar to routine work undertaken in forensic laboratories. Like SNPs, they have short amplicons that make them suitable for the profiling of degraded DNA samples (Borsting *et al.* 2013).

The major limitations for the genetic tests to derive ancestry information include the lack of sufficiently population data and difficulty in the assessment of complex admixture patterns in individuals with co-ancestry (Willuweit and Roewer 2007; King *et al.* 2007; Wetton *et al.* 2005).

4.1.1. Objectives and Aims

The aim of this Chapter was to estimate the effectiveness of INDELs and STRs to predict ancestry information of five tested populations (Qatari, Pakistani, Sudanese, Tunisian and Yemeni). To achieve this, following objectives were set.

(a) To test the population differentiation by using pairwise F_{ST} values for five tested populations of Qatar.

(b) To assess the population sub-structure using the cluster analysis for five tested populations of Qatar.

(c) To calculate the likelihood ratio for the assignment of five tested populations of Qatar.

(d) To compare the efficiency of INDELs and STRs to estimate the ancestry information for five tested populations of Qatar.

4.2. Materials and Methods

This Chapter involved statistical analysis of the data by employing different software. In this study, population differentiation was tested using pairwise F_{ST} values through Arlequin v3.1 software (Excoffier *et al.* 2010), population assignment of individual genotypes was done by calculating individual maximum likelihood through Snipper v2.0 software (Phillips *et al.* 2007) and assessment of predictive value of the marker set by cluster analysis through STRUCTURE v2.3.4 software (Pritchard *et al.* 2010).

4.3. RESULTS

4.3.1 Estimation of geographical origin using INDELs

4.3.1.1 Population Genetic Structure

In order to assess the population differentiation among five populations based in Qatar using 30 INDELs, pairwise F_{ST} values were determined using Arlequin software (Table 4.1). The greatest pairwise F_{ST} value was observed between Pakistani (South Asian) and Sudanese (East Africa) populations at 3% and lowest pairwise F_{ST} of 0.052% was recorded between Qatari and Yemeni populations (Middle Eastern) indicating their closer relationship in terms of origin.

Table 4.1 Table showing the pairwise F_{ST} values among 5 tested population based on 30 INDELs (*result not significant).

	Yemeni	Tunisian	Sudanese	Qatari	Pakistani
Yemeni	0				
Tunisian	0.00553	0			
Sudanese	0.02201	0.00892	0		
Qatari	**0.00052***	0.00359	0.01511	0	
Pakistani	0.01356	0.01673	**0.03112**	0.0108	0

4.3.1.2 Cluster Analysis

This study further assessed the population sub-structure using the STRUCTURE 2.3.4 software (Pritchard *et al.* 2010). This programme is based on Bayesian method and assigns the samples to K populations or clusters on the basis of allele frequencies. Various population models can be assumed before analysis in this software. All of five populations in this research were analysed using admixture model with correlated allele frequencies i.e. individuals may have some common alleles based on genetic drift, thus sharing ancestry (Rosenberg *et al.* 2005).

The software allows setting the number of populations (k) prior to analysis. However, during this study, it was not done and the software was directed to assess the structure of populations based on the values of K from 1 to 5.

In order to estimate the value of K (number of populations), the output of result file (zip format) from STRUCTURE software was loaded onto web based STRUCTURE Harvester (http://taylor0.biology.ucla.edu/structureHarvester). This software use Evanno approach based on the values of the rate of change in the log probability of data between successive K values (ΔK).

The results described in Table 4.2 and Figure 4.1 have shown that the calculated value of K by using STRUCTURE Harvester with Evanno approach (Evanno *et al.* 2005), was found to be 2, which indicated that two populations were the most likely, based on the INDEL data.

Table 4.2 Table showing the Evanno Table containing likelihood values generated from STRUCTURE-Harvester software for 5 tested populations using 30 INDELs data.

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	10	-20024.110000	0.242441	—	—	—
2	10	-20069.440000	8.440405	-45.330000	519.860000	61.591828
3	10	-20634.630000	75.687149	-565.190000	273.610000	3.615013
4	10	-20926.210000	249.493133	-291.580000	170.750000	0.684388
5	10	-21047.040000	261.540760	-120.830000	—	—

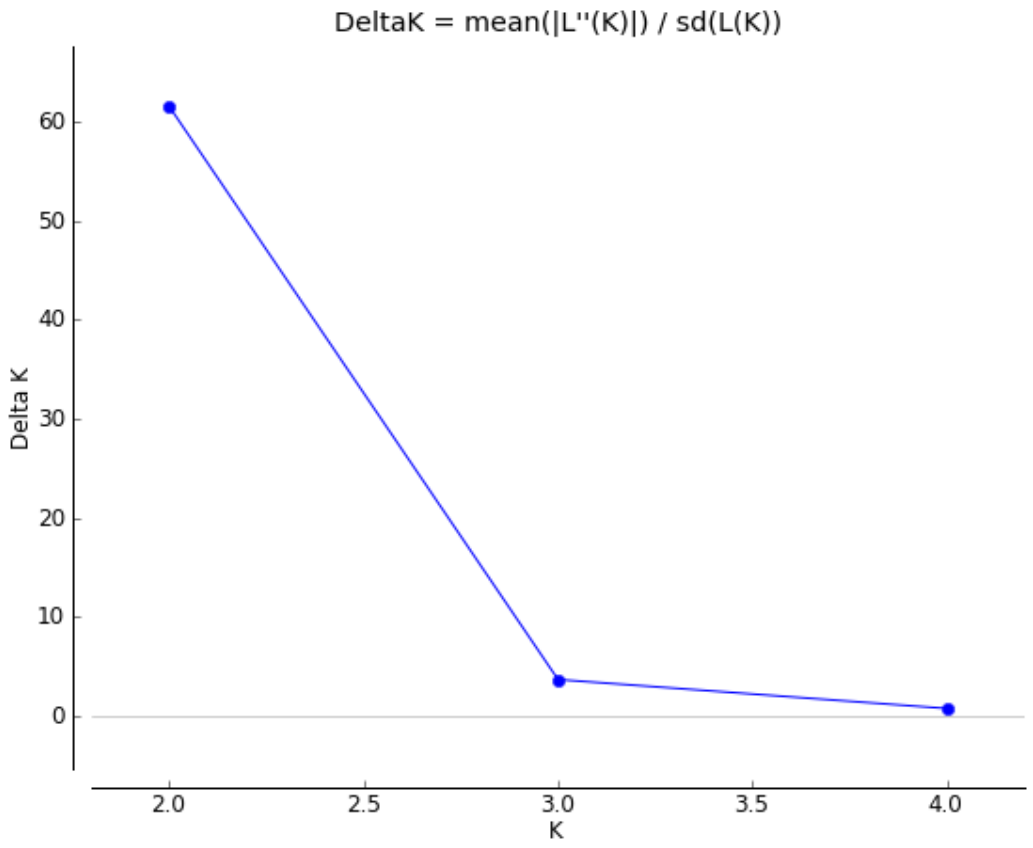

Figure 4.1 Figure showing the graphical representation of assigning K value based on Evanno method generated from STRUCTURE-Harvester software for 5 tested populations (n=5) using data generated from 30 INDELs.

4.3.1.3 Population assignment of individual genotypes

The likelihood of individual INDELs genotypes belonging to each of five population groups (Qatari, Pakistani, Sudanese, Tunisian and Yemeni) was calculated using Snipper software. In order to analyse INDEL data on the Snipper software (originally developed for SNPs analysis), the INDEL profiles were designated as A/C SNPs with deletion alleles as "A" and insertion alleles as "C". Similarly, in order to predict the likelihood calculation for the INDELs, a cross validation was performed using the option of "Thorough analysis of population data" with verbose cross-validation analysis using the best 30 SNPs. Each sample was tested as unknown against the training set of all other samples. The results of success ratio using different combinations of populations are shown in Tables 4.3-4.10 and Figures 4.2- 4.9.

While performing thorough analysis with verbose cross-validation for five population groups (Table 4.3 and Figure 4.2), after applying Hardy-Weinberg Equilibrium (HWE), the results gave low success ratios for assigning populations (except Tunisian at 93%).

Then, a thorough analysis with verbose cross-validation after applying HWE was performed for four populations by excluding Pakistani population due to the fact that it was the most distant population (Table 4.4 and Figure 4.3). The results indicated that Tunisian population had the highest likelihood ratio (95%). While the other three populations showed non-significant percentages (Qatari population with lowest value of 9%).

Based on the pairwise F_{ST} values, when three populations (out of five) were tested in different combinations with each other, the success ratio of each population was increased to more than 90% in different combinations (except the Qatari population which remain lowest in all of the combinations). These results are displayed in Tables 4.5-4.9 and Figures 4.4-4.8.

Table 4.3 Table showing the population assignments generated from Snipper 2.0 software by using cross verbose-validation after applying HWE for 5 tested populations based on 30 INDELs.

	Qatar	Pakistani	Sudanese	Yemeni	Tunisian
Population of Qatari origin	7.00 %	19.00 %	17.00 %	32.00 %	25.00 %
Population of Pakistani origin	0.00 %	25.00 %	15.00 %	37.00 %	23.00 %
Population of Sudanese origin	0.00 %	0.00 %	33.00 %	16.00 %	51.00 %
Population of Yemeni origin	0.00 %	1.00 %	1.00 %	30.00 %	68.00 %
Population of Tunisian origin	1.00 %	2.00 %	2.00 %	2.00 %	93.00 %

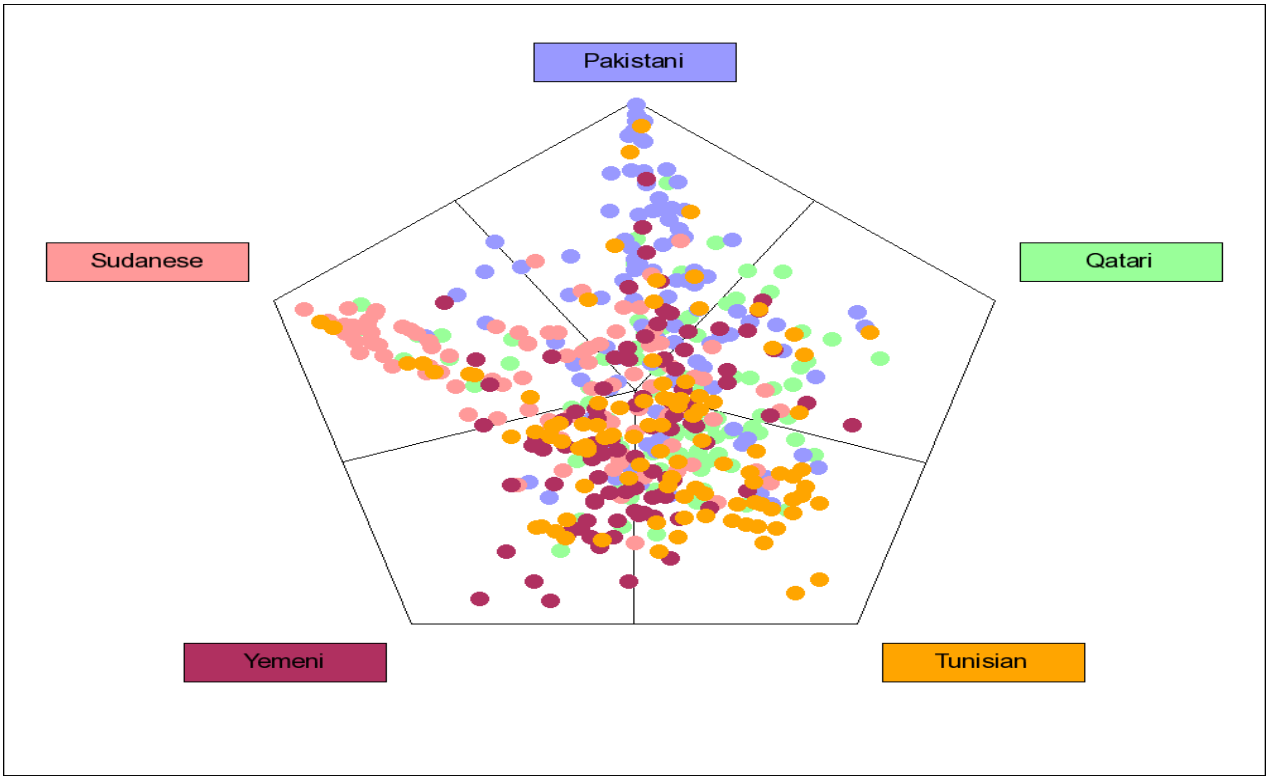

Figure 4.2 Figure showing the 2-D model of the success ratio of population assignments generated from Snipper 2.0 software for 5 tested populations based on 30 INDELs. It is evident from the above diagram that most of the populations are concentrated in the middle of pentagon, which indicates that they cannot be differentiated by using INDELs.

Table 4.4 Table showing the population assignments generated from Snipper 2.0 software by using cross verbose-validation after applying HWE for 4 tested populations based on 30 INDELs.

	Qatari	Sudanese	Yemeni	Tunisian
Population of Qatari origin	9.00 %	20.00 %	39.00 %	32.00 %
Population of Sudanese origin	0.00 %	33.00 %	16.00 %	51.00 %
Population of Yemeni origin	0.00 %	1.00 %	31.00 %	68.00 %
Population of Tunisian origin	1.00 %	2.00 %	2.00 %	95.00 %

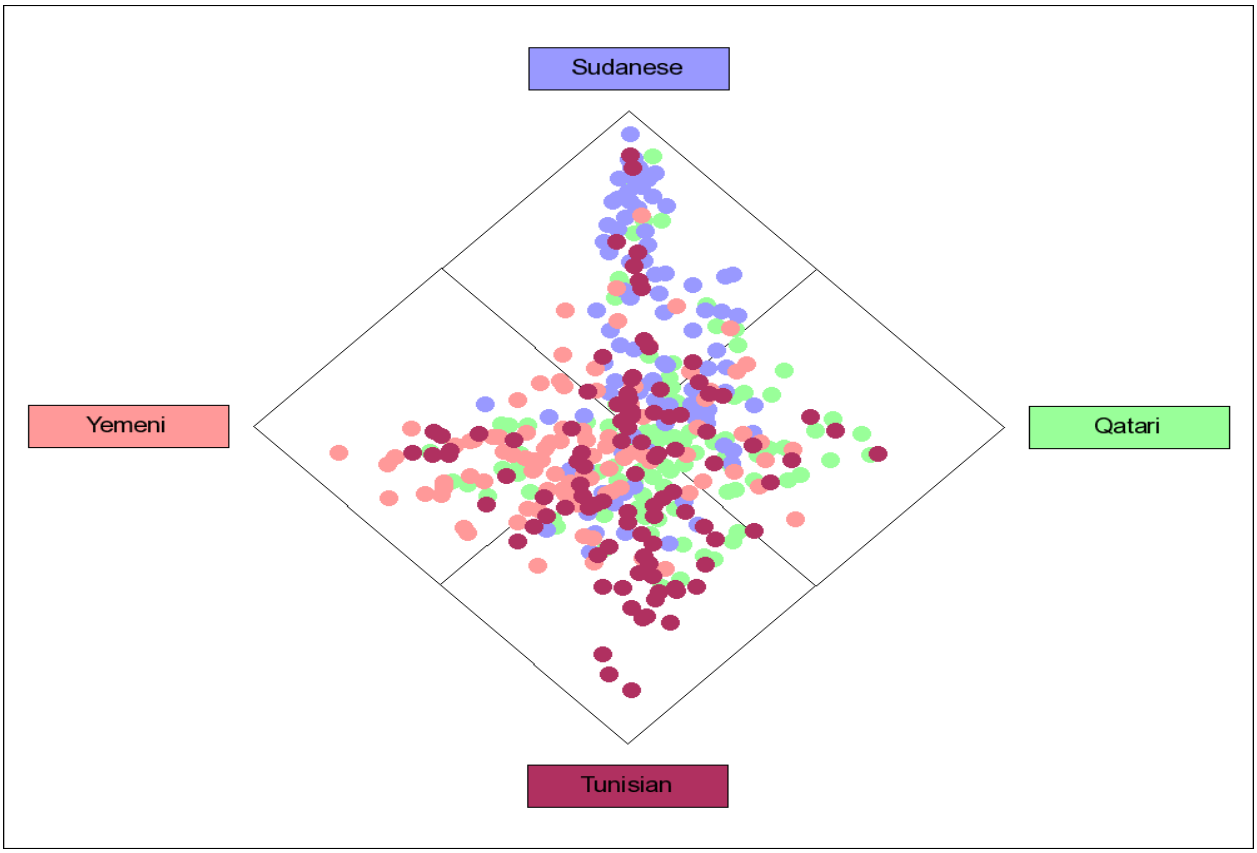

Figure 4.3 Figure showing the 2-D model of the success ratio of population assignments generated from Snipper 2.0 software for 4 tested populations based on 30 INDELs. It is evident from the above diagram that all of the four populations are concentrated in the middle of pentagon rather than on its margin which indicates that they cannot be differentiated by using INDELs.

Table 4.5 Table showing the population assignments generated from Snipper 2.0 software by using cross verbose-validation after applying HWE for 3 tested populations based on 30 INDELs.

	Qatari	Pakistani	Sudanese
Population of Qatari origin	17.00 %	46.00 %	37.00 %
Population of Pakistani origin	0.00 %	51.00 %	49.00 %
Population of Sudanese origin	0.00 %	2.00 %	98.00 %

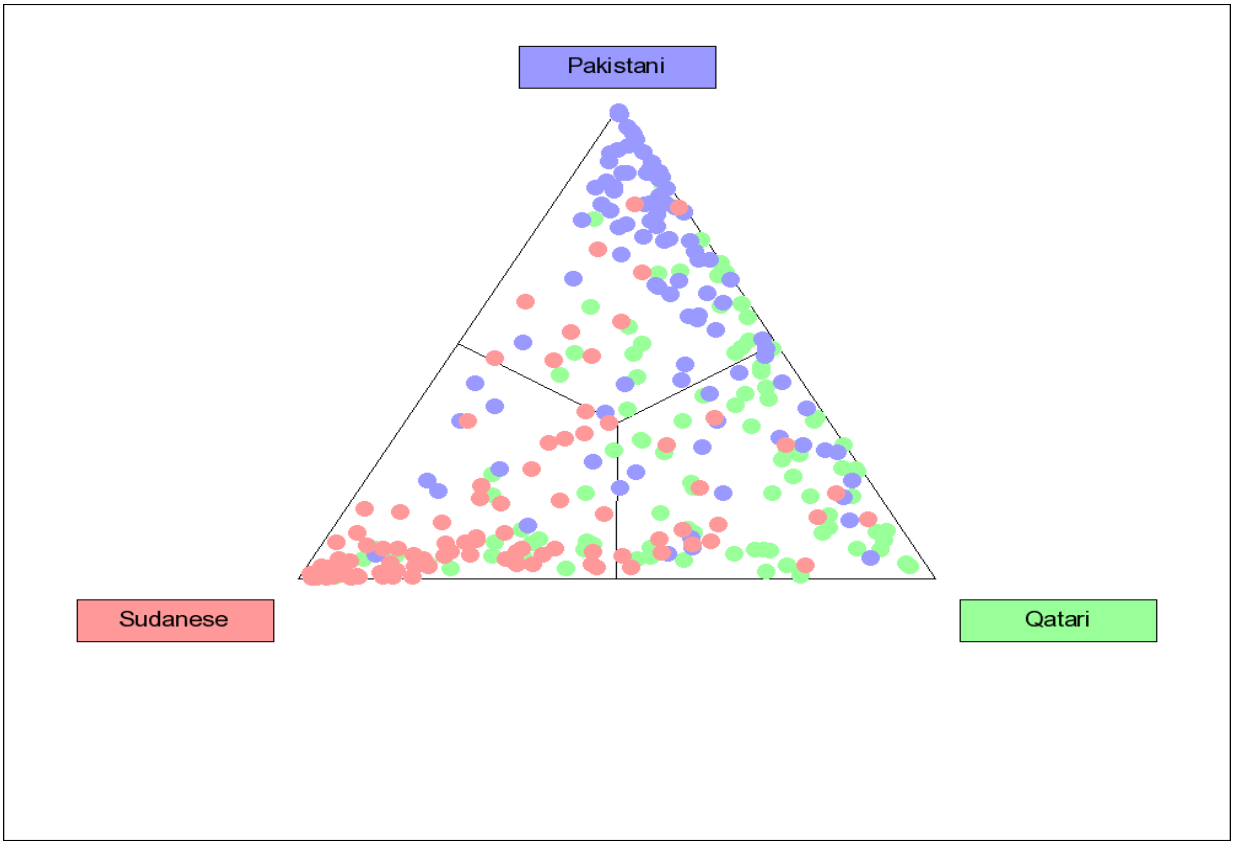

Figure 4.4 Figure showing the 2-D model of the success ratio of population assignments generated from Snipper 2.0 software for 3 tested populations based on 30 INDELs. It is evident from the above diagram that most of the Snipper data from the populations of Sudanese and Pakistani are concentrated on the margin, indicating that they are better differentiated than that of Qatari population.

Table 4.6 Table showing the population assignments generated from Snipper 2.0 software by using cross verbose-validation after applying HWE for 3 tested populations based on 30 INDELs.

	Pakistani	Sudanese	Yemeni
Population of Pakistani origin	27.00 %	22.00 %	51.00 %
Population of Sudanese origin	0.00 %	45.00 %	55.00 %
Population of Yemeni origin	4.00 %	2.00 %	**94.00 %**

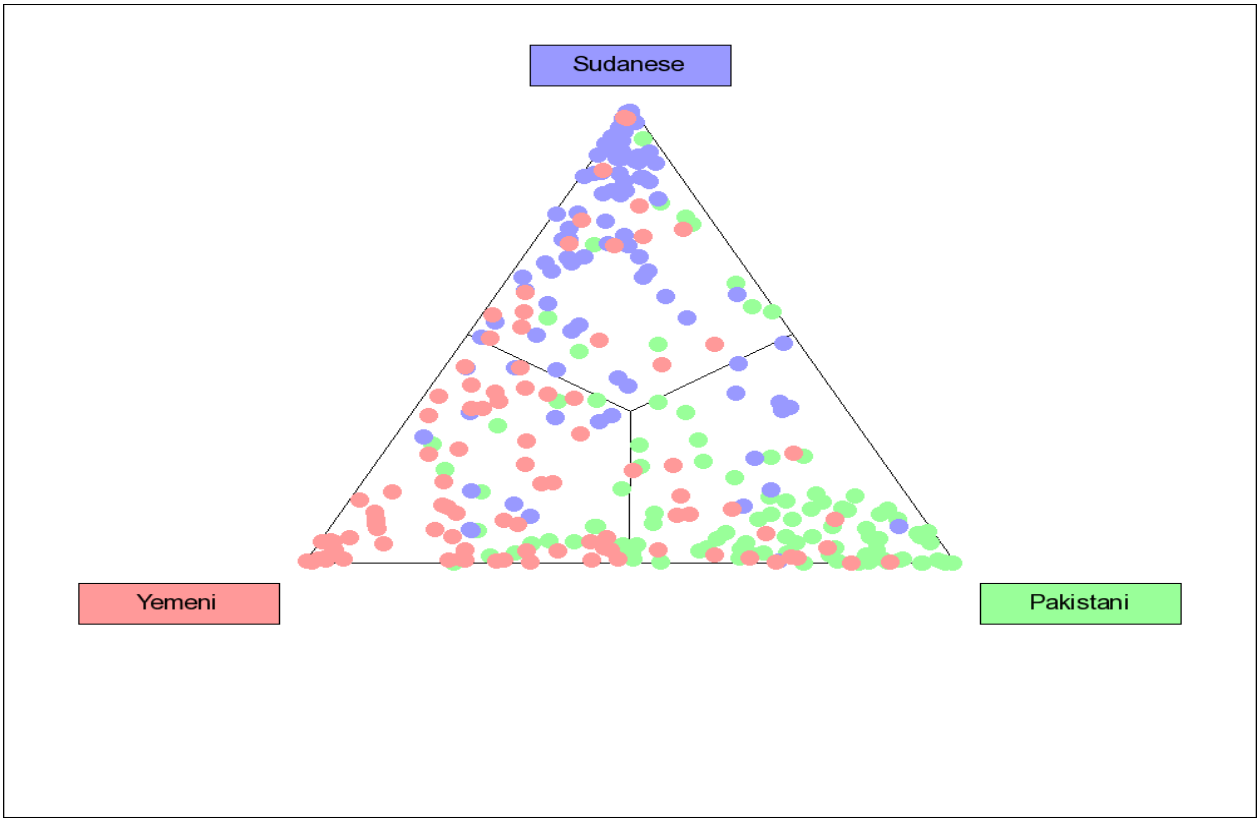

Figure 4.5 shows the 2-D model of the success ratio of population assignments generated from Snipper 2.0. Software for 3 tested populations based on 30 INDELs. It is evident from the above diagram that most of the Snipper data from Yemeni population are concentrated in its respective box, indicating that it can be better differentiated than that of Sudanese and Pakistanis.

Table 4.7 Table showing the population assignments generated from Snipper 2.0 software by using cross verbose-validation after applying HWE for 3 tested populations based on 30 INDELs.

	Qatari	Sudanese	Tunisian
Population of Qatari origin	19.00 %	25.00 %	56.00 %
Population of Sudanese origin	0.00 %	34.00 %	66.00 %
Population of Tunisian origin	2.00 %	2.00 %	**96.00 %**

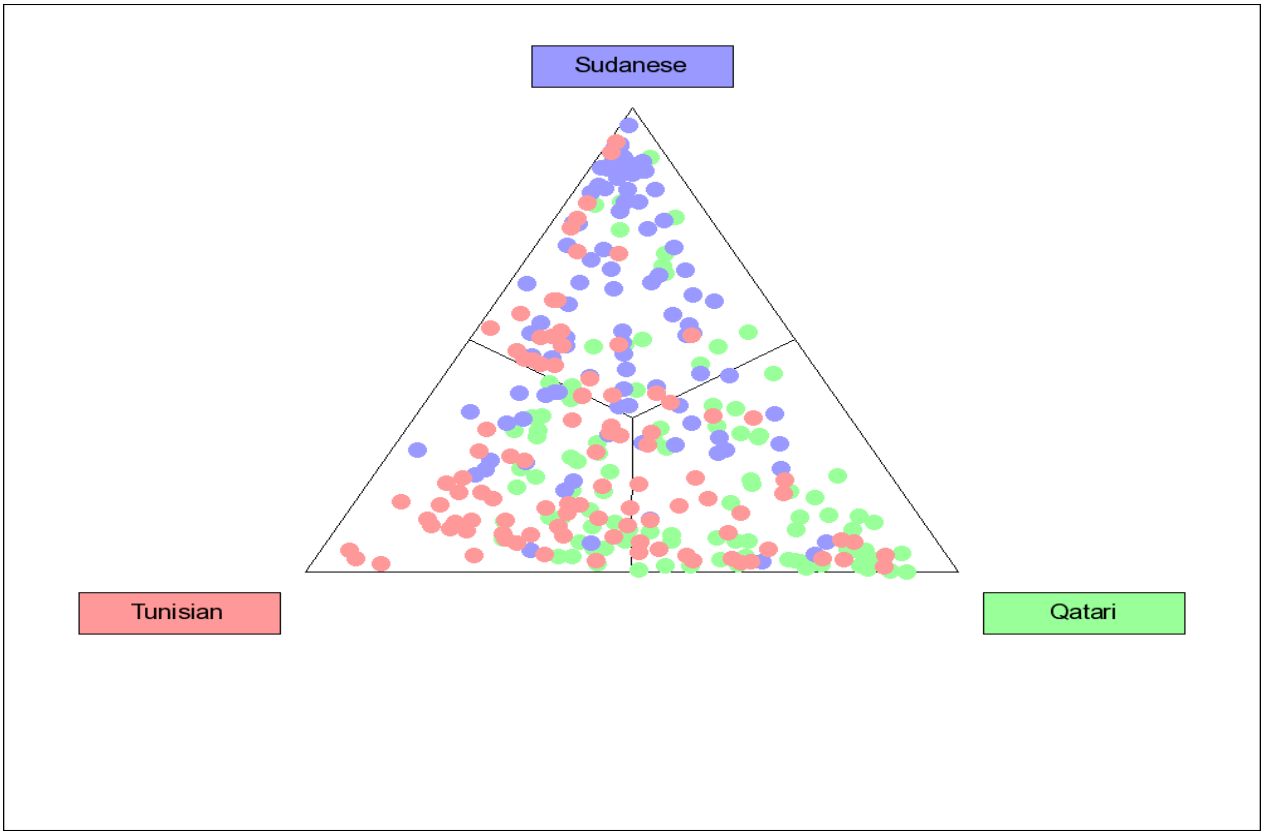

Figure 4.6 Figure showing the 2-D model of the success ratio of population assignments generated from Snipper 2.0. Software for 3 tested populations based on 30 INDELs. It is evident from the above diagram that most of the Snipper data from Tunisian population are concentrated in its respective box, indicating that it can be better differentiated than that of Sudanese and Qataris.

Table 4.8 Table showing the population assignments generated from Snipper 2.0. Software by using cross verbose-validation after applying HWE for 3 tested populations based on 30 INDELs.

	Qatari	Yemeni	Tunisian
Population of Qatari origin	13.00 %	47.00 %	40.00 %
Population of Yemeni origin	1.00 %	31.00 %	68.00 %
Population of Tunisian origin	1.00 %	2.00 %	**97.00 %**

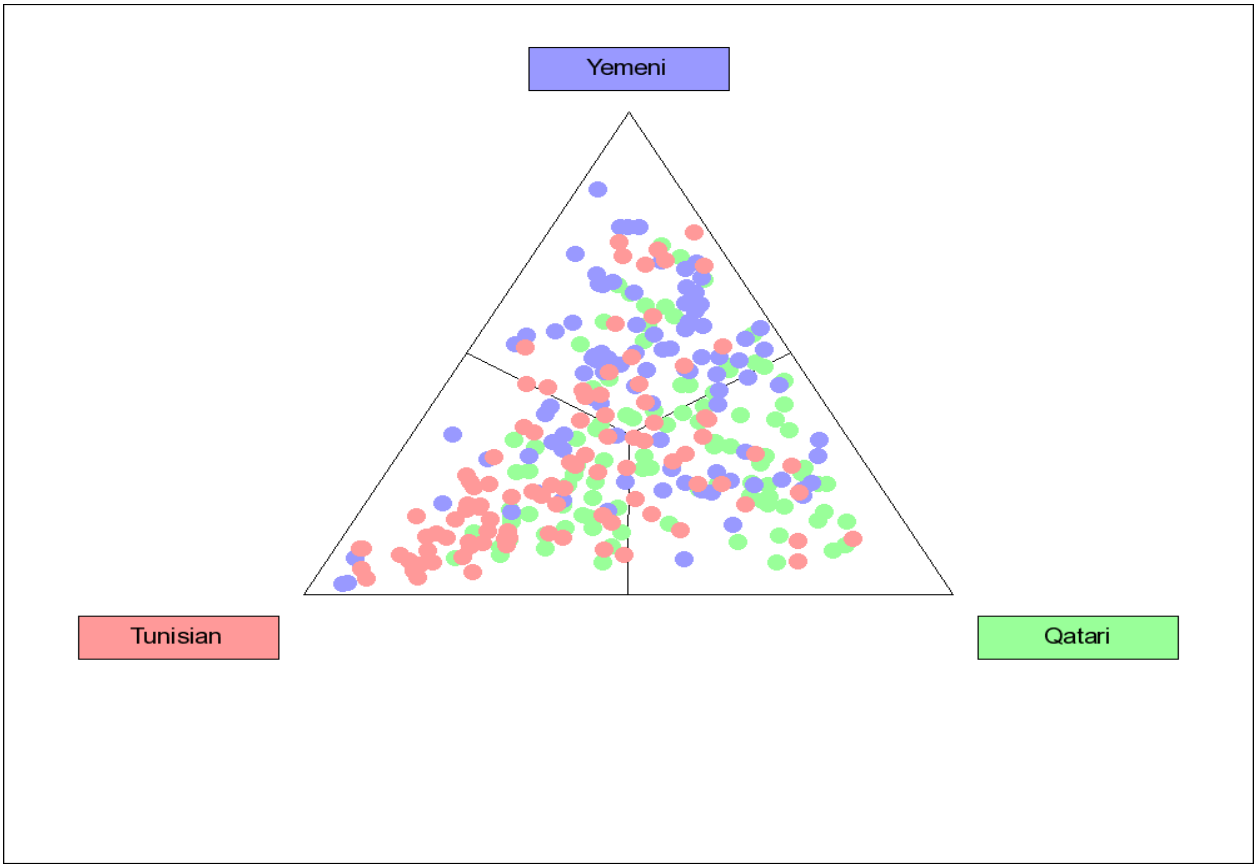

Figure 4.7 Figure showing the 2-D model of the success ratio of population assignments generated from Snipper 2.0. Software for 3 tested populations based on 30 INDELs. It is evident from the above diagram that most of the Snipper data from Tunisian population are concentrated in its respective box, indicating that it can be better differentiated than that of Yemenis and Qataris.

Table 4.9 Table showing the population assignments generated from Snipper 2.0. Software by using cross verbose-validation after applying HWE for 3 tested populations based on 30 INDELs.

	Sudanese	Tunisian	Pakistani
Population of Sudanese origin	30.00 %	52.00 %	18.00 %
Population of Tunisian origin	0.00 %	38.00 %	62.00 %
Population of Pakistani origin	2.00 %	6.00 %	**92.00 %**

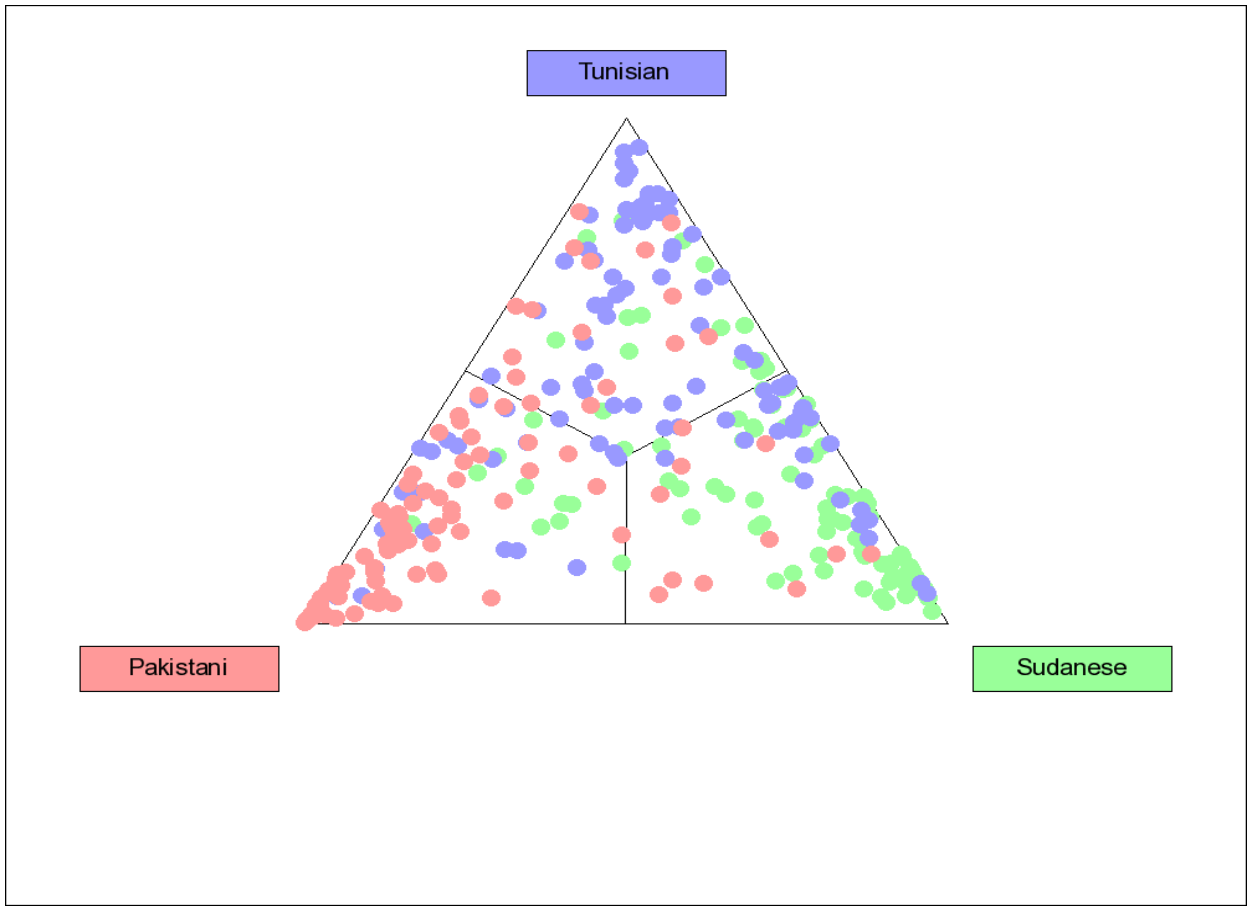

Figure 4.8 Figure showing the 2-D model of the success ratio of population assignments generated from Snipper 2.0. Software for 3 tested populations based on 30 INDELs. It is evident from the above diagram that most of the Snipper data from Pakistani population are concentrated in its respective box, indicating that it can be better differentiated than that of Sudanese and Tunisians.

4.3.2 Estimation of geographical origin using STRs

4.3.2.1 Pairwise F_{ST} values

Table 4.10 shows the pairwise F_{ST} values between five populations based in Qatar using STR data. The comparison of pairwise F_{ST} values using 15 STRs among 5 tested populations confirmed the same results as obtained using 30 INDELs that Qatari and Yemeni had the lowest pairwise F_{ST} value (0.15%) revealing their close relatedness due to close geographical positions. The Pakistani population showed the highest pairwise F_{ST} values (upto 1.2%) in any relation to other populations which indicated the greatest differentiation from the other population groups.

Table 4.10 Table showing pairwise F_{ST} values between 5 tested population based on 15 STRs (*result not significant).

	Yemeni	Tunisian	Sudanese	Qatari	Pakistani
Yemeni	0				
Tunisian	0.00662	0			
Sudanese	0.01087	0.00451	0		
Qatari	**0.00155***	0.00371	0.00851	0	
Pakistani	0.00850	0.00810	**0.01199**	0.00745	0

4.3.2.2 Cluster Analysis

The parameters of STRUCTURE software to analyse 15 STRs genetic data were similar as were used for INDELs analysis. The calculation of K value for STRs data was found to be 4 rather than 2 (for INDELs). The results of STRUCTURE Harvester for STRs are shown in Table 4.11 and Figure 4.9.

Table 4.11 Table showing the Evanno likelihood values generated from STRUCTURE-Harvester software for 5 tested populations using 15 STRs data.

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	10	-27338.200000	0.216025	—	—	—
2	10	-27857.230000	218.410724	-519.030000	436.790000	1.999856
3	10	-27939.470000	257.286775	-82.240000	241.240000	0.937631
4	10	-28262.950000	237.960605	-323.480000	1203.330000	5.056845
5	10	-29789.760000	709.627159	-1526.810000	—	—

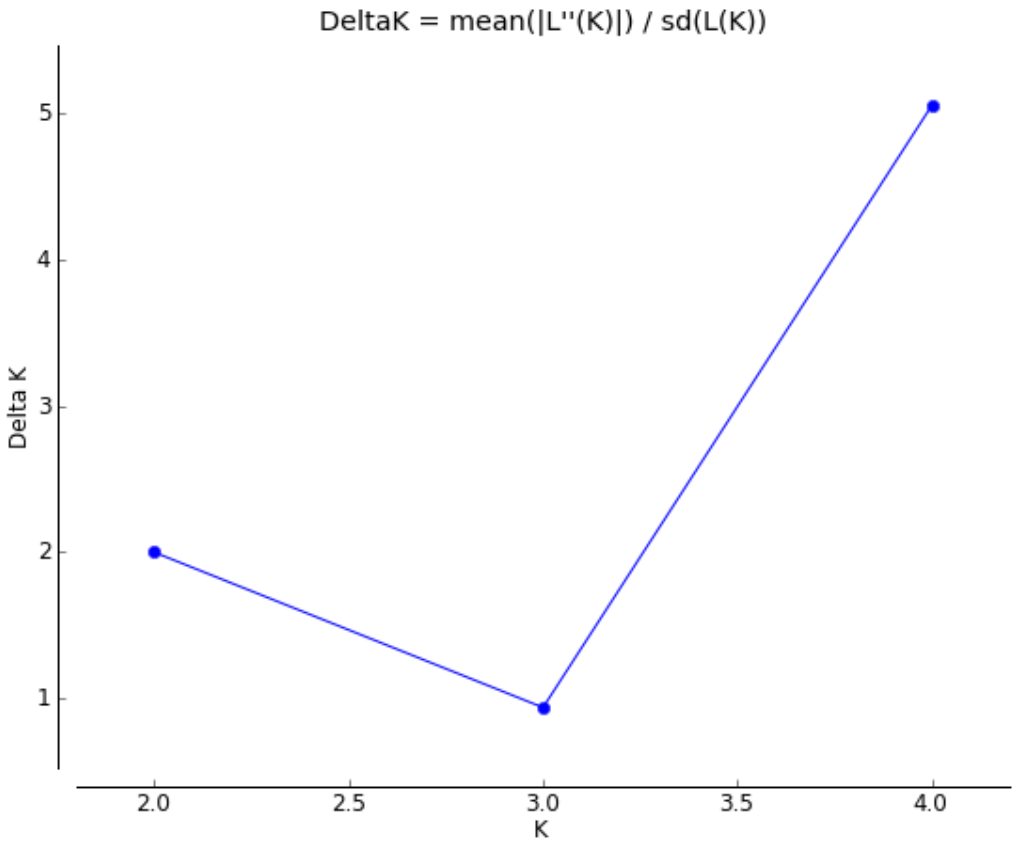

Figure 4.9 Figure showing the graphical representation of assigning K value based on Evanno method generated from STRUCTURE-Harvester Software for 5 tested populations (n=5) using 15 STRs data.

4.3.2.3 Snipper Analysis

The online portal of Snipper App *suite* v2.0 is made for SNPs, but it also provides an option to analyse STRs. This option allows an unknown STR profile to compare to a set of reference populations (training set of STRs data).

The STR data from five populations were arranged according to template Excel file and then this file was loaded onto the option of "Classification with 32 STR training set or a custom Excel file of frequencies" along with mentioning of the number of populations tested. Then, the unknown STR profile was typed and compared against the reference populations. Based on the likelihood calculation, it predicted the population along with percentages. Table 4.12 shows the outputs of some profiles with known nationalities.

Table 4.12 Table showing the assignment of nationalities based on likelihoods of some profiles with known nationalities.

	likelihoods				
PROFILE FROM	**QATARI**	**PAKISTANI**	**SUDANESE**	**YEMENI**	**TUNISIAN**
QATARI	**7.11134e-17**	6.78181e-17	8.54639e-19	8.38716e-19	9.38774e-18
	47.40 % for QATARI; 45.21 % for PAKISTANI; 6.26 % for TUNISIAN.				
PAKISTANI	6.42771e-19	**4.48445e-18**	5.22387e-20	7.74352e-19	2.37800e-18
	53.82 % for PAKISTANI; 28.54 % for TUNISIAN; 9.29 % for YEMENI.				
SUDANESE	2.78965e-20	4.66173e-20	**1.21046e-19**	6.27208e-21	1.83913e-20
	54.97 % for SUDANESE; 21.17 % for PAKISTANI; 12.67 % for QATARI.				
YEMENI	6.28636e-16	1.92579e-16	2.01701e-16	**2.90561e-15**	2.17606e-16
	70.08 % for YEMENI; 15.16 % for QATARI; 5.25 % for TUNISIAN.				
TUNISIAN	2.75286e-19	8.81750e-19	2.54474e-19	2.98572e-19	**1.96905e-18**
	53.52 % for TUNISIAN; 23.97 % for PAKISTANI; 8.12 % for YEMENI.				

4.4 Discussion

In this Chapter, an effort has been made to derive the ancestry information by analysing and comparing data of STRs and INDELs generated from 500 samples belonging to five different populations based in Qatar. During this study, the sampling from distant populations helped greatly to understand the population differentiation due to their different geographical positions.

Pairwise F_{ST} values for 5 tested populations revealed that in most of the pairs of populations, pairwise F_{ST} values were high which indicated genetic differentiation among them, according to their geographical locations (Monica *et al.* 2005; Ramachandran *et al.* 2005). The lowest F_{ST} value (0.052%) for Qatari-Yemeni combination showed their close relatedness due to possible common ancestors i.e. nomads from Arabian Peninsula (Carol *et al.* 2001). Another evidence of their close relatedness can be supported by a Y-chromosome study, which demonstrated that Qatari and Yemeni shared most of Y haplogroups (Cadenas *et al.* 2008).

The results calculated from Snipper analysis of INDELs showed that by using different combinations of populations, the four populations (Pakistani, Sudanese, Tunisian and Yemeni) were assigned correctly (> 90%) according to their origins. The only exception was Qatari population, which did not show significant likelihood in any of tested combinations.

In a comprehensive population study (Carol *et al.* 2001), it was investigated that the possible ethnic groups from which Qataris were originated, could be three; Bedouins (descendants of Arab nomads), Hadar or Irani-Qataris (migrated from Iran, Pakistan and Afghanistan) and Abd (slaves brought from East Africa mainly from Sudan and Somalia).

In addition, the population studies based on mitochondrial DNA, revealed that more than 35% of the Qatari lineages were from African ancestry (East African and Sub-Saharan) and the rest of the lineages being Eurasian (Rowold *et al.* 2007).

It can be concluded that current Qatari population is composed of not only native Bedouins, but also contain African and Asian ancestors.

The STRUCTURE analysis using INDELs data indicated that there was limited genetic differentiation among the populations as the value of K was found to be 2 rather than 5, according to number of populations tested. The maximum likelihood or ΔK calculated by Evanno approach (Evanno *et al*. 2005) from an average of all 10 runs, was observed, when K = 2. On the other hand, the STRUCTURE results for STRs data indicated the presence of four populations i.e. K = 4. The different outcomes of STRUCTURE analysis for INDELs and STRs can be explained in terms of their polymorphic nature of markers. STR markers are more polymorphic (Gill 2002) as compared to INDEL markers, hence the value of K (the number of populations derived) for STRs was observed more than that of INDELs and also closer to accurate number (5) of tested populations. In addition to polymorphic nature of STRs, the pairwise F_{ST} values generated in this research and Y-chromosome study carried out for Middle Eastern populations (Cadenas *et al.* 2008) indicated that Qatari and Yemeni populations are closely related to each other, so if these two populations are considered as single population, then the value of K derived from STR data is not surprising.

Similar efforts were made to assess the ability of Identifiler Plus® (Applied Biosystems™) based 15 STRs for ancestry-informativeness (Londin *et al.* 2010; Phillips *et al.* 2011) but failed to differentiate the global sample set of 7 populations. However when 36 novel STRs, including 33 dinucleotide-repeat STRs, were tried, better ancestry information was obtained (Pereira *et al.* 2012). To derive ancestry information, di-nucleotide repeat STR loci were found better than those of tri- or tetra-nucleotide repeat loci because they were more highly differentiated across populations. Although di-nucleotide STRs cannot be easily used for forensic identification purpose due to their high stutter levels (usually 30%) and difficulty in resolving size based electrophoretic separation (Walsh *et al.* 1996), they are better when attempting to infer ancestry than tetra-nucleotide repeat STRs.

A total of 30 autosomal INDELs were analysed in this study, which were developed for identification purposes rather than ancestry inference. Studies were performed in which AIM-INDEL panels were developed for ancestry informativeness which include 48 INDELs in three multiplexes form (Santos *et al.* 2010), 48 INDELs as single multiplex (Pereira *et al.* 2012), and 21

INDELs as single multiplex (Zaumsegel *et al.* 2013). These AIM-INDELs provide a simple tool to derive ancestry information from a single test (Pereira *et al.* 2009). The best genetic markers which have been proved to derive ancestry information are SNPs. Along with other PCR assays, 34-plex and 47-plex SNP assays have been developed using ancestry informative SNPs (Phillips *et al.* 2007; Kersbergen *et al.* 2009; Fondevila *et al.* 2013).

However, how any of these assays would perform using the samples studied in this research is not clear. But as the populations are not highly differentiated, they may not be so effective in the context of Qatar.

4.5. Conclusion

The genetic data generated for INDELs and STRs during this study were originally meant to evaluate their forensic efficiency but in this Chapter, an attempt was made to derive ancestry information. After analyzing the genetic data from INDELs and STRs, it can be concluded that autosomal INDELs cannot differentiate the populations accurately. They can only be used for identity purposes in forensic DNA analysis.

CHAPTER 5
DEVELOPMENT, OPTIMISATION AND VALIDATION OF A NEW MULTIPLEX (MINI-INDELS) PCR ASSAY

5.1. OVERVIEW

Recovery of genetic information from a poor quality of DNA sample is a challenging task using conventional short tandem repeat (STR) analysis. In many of the cases, the results obtained using commercial STR kits results in a partial or no DNA profile due to fragmentation of DNA (Whitaker *et al.* 1995; Li *et al.* 2011; Manta *et al.* 2013). As an alternative to STRs, some laboratories have opted to single nucleotide polymorphism (SNP) analysis to obtain more information. This approach, while helpful, requires a large number of SNPs to be typed to have same discriminating power as STRs, which is time consuming and costly (Gill *et al.* 2001). There is a third approach which is emerging as an alternative tool for the recovery of DNA profiles from challenged samples and this is the Insertion Deletion (INDEL) polymorphism, which is due to the presence of some of the combined advantages of STRs and SNPs; wide distribution in human genome, lower mutation rates and use of smaller amplicons and routine genotyping techniques (Li *et al.* 2011; Manta *et al.* 2013). These characteristics make them ideal candidates for forensic applications, especially for the profiling of degraded and inhibited samples (Romanini *et al.* 2012).

Several INDEL multiplexes have been developed including several in-house (Phillips *et al*. 2007; Pereira *et al.* 2009; Fondevila *et al.* 2011; Kis *et al.* 2012) and one commercial kit for autosomal INDELs (Qiagen™ Investigator® DIPplex kit). This kit was commercially launched in 2009 and it allows simultaneous amplification of 30 biallelic Deletion/Insertion Polymorphisms (DIPs), which are spread over 19 autosomes and are separated by at least 10 Mbp from commercially available STRs. A specific fragment of the amelogenin gene is also typed as a sex marker. The

maximum amplicon length of allele in the DIPplex kit is 160 bp (Qiagen™ Investigator® DIPplex handbook).

This study developed an amplification kit of a multiplex of mini-INDELs by re-designing 14 of larger amplicons from the Qiagen DIPplex PCR Amplification kit: HLD67, HLD131, HLD6, HLD101, HLD124, HLD39, HLD58, HLD88, HLD99, HLD84, HLD97, HLD125, HLD128, HLD111 and the sex identification locus amelogenin.

5.1.1 Objectives

The objective of this Chapter was to develop a mini-INDELs based multiplex which included:

(g) To develop a multiplex PCR assay with short amplicons.

(h) To do concordance study between Qiagen™ Investigator® DIPplex kit and the in-house developed mini-INDELs kit.

(i) To optimize the mini-INDELs multiplex assessing reproducibility, sensitivity and performance using high molecular weight and challenged DNA samples.

(j) To calculate the discriminatory power of the mini-INDELs markers.

5.2 MATERIALS AND METHODS

5.2.1 Marker selection

The markers for this study were chosen from the INDEL markers from Qiagen™ Investigator® DIPplex kit. The basic criteria of the markers selection were their larger amplicon size. A total of 20 Markers were initially selected which included HLD6, HLD39, HLD40, HLD56, HLD58, HLD64, HLD67, HLD70, HLD81, HLD84, HLD88, HLD97, HLD99, HLD101, HLD111, HLD122, HLD124, HLD125, HLD128 and HLD 131. The sequences of these markers were obtained using their SNP ID# through NCBI website. The sequence of two of markers, HLD40 and HLD70, were not available on NCBI website.

5.2.2 Alignment of the INDEL sequences

BioEdit Sequence Alignment Editor (Hall 2011) is a tool that is used to edit, align, manipulate and analyze the protein and nucleic acid sequences (www.mbio.ncsu.edu/bioedit). The purpose to use BioEdit tool in this study was to align the downloaded sequences from different ethnic groups/nationalities, so that any difference in the flanking region could be identified.

In order to achieve this goal, the Fasta sequence of each INDEL marker available on NCBI website for different ethnic groups were copied and then pasted into New Sequence tab of BioEdit software. In this way, the Fasta sequences from different available population groups were aligned together. The population groups from which sequences were aligned, included East Asians (Koreans, Japanese), African North Americans (African, Amerind), and European (Irish). The sequences of some population groups were done more than once in different studies.

5.2.3 Miniplex (mini-INDELs) approach to DNA degradation

The concept of miniplexes (reduced size primer sets) were introduced to address the issue of DNA degradation. A conventional profile from a degraded sample showed the loss of high base pair loci in sloping pattern (Figure 5.1). For the analysis of degraded samples, a new set of primers (mini-plex) could be developed in which primers of particular markers could be re-designed to produce the smaller amplicons (Butler *et al.* 2003).

In this study, the primers for markers having higher amplified product size (ranging from 120 bp to 160 bp) present in Investigator® DIPplex kit, were re-designed as mini-INDELs by decreasing their product sizes (Table.5.1).

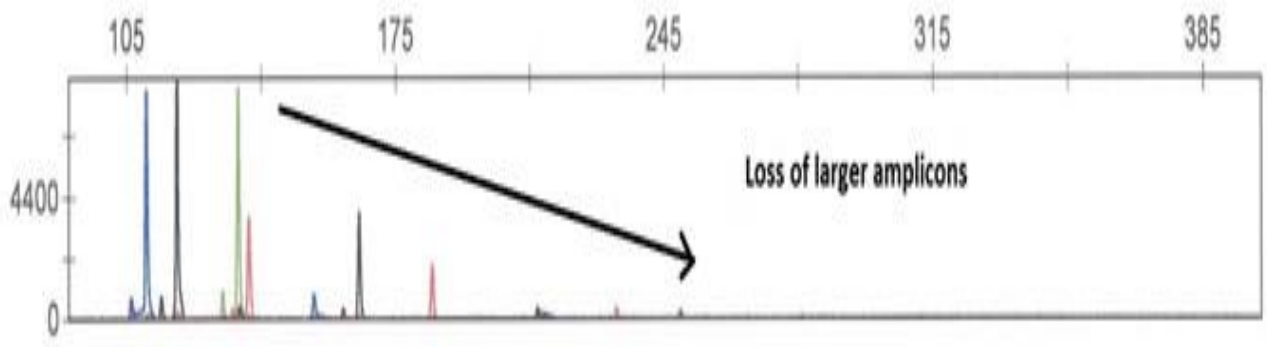

Figure 5.1 Figure showing a profile using Applied Biosystems™ Identifiler Plus® kit showing the loss of larger amplicons due to DNA degradation.

5.2.4. Multiplex INDEL primer Design

In this project, the primer design of 18 INDEL markers was attempted using the Primer- blast software from NCBI website by applying the following criteria for the PCR primers;

Amplicon size = 75-130 bp

Tm = 55-60 °C

GC content = 45-55 %

Subsequently, the primer pairs obtained were checked for hairpin and primer-dimer secondary structures using OligoAnalyzer 3.1 software (www.idtdna.com). Fourteen markers were selected after going through all the above criteria.

5.2.5. Assigning fluorescent labelled dyes

All the primers were then organized for expected amplicon size and assigned to four different fluorescent labeled dyes from life technology (FAM, NED, VIC and PET). The primers were distributed among four dyes to achieve a distance of at least 2 base pair between any two alleles. A tail of 5 bases (TGAAT) was added to HLD67 marker to achieve an even distance from neighboring alleles (Table 5.2).

5.2.6. Primer synthesis and purity

Primer pairs with 5' fluorescein labelled forward primers and unlabelled reverse primers were synthesized (Applied Biosystems™, UK) purified using HPLC and desalting and delivered in lyophilised.

Each individual primer was synthesized at 10 nM scale and a stock solution was prepared for each primer by dissolving 100 µl molecular grade distilled water to achieve the concentration of 100 µM. Stock solutions were stored at -20 °C.

Table 5.1 Table showing the re-designing of Mini-INDELs primers from DIPplex kit showing their gene accession #, chromosome localization, motif, Amplicon sizes of both kits and decrease in size for mini-INDELs primers.

Locus	Gene accession #	Chromosome localization	Motif(+DIP)	Amplicon expected size (bp)	Amplicon size in DIPplex kit (bp)	Decrease in size (bp)
HLD6	rs1610905	16q13	GCAGGACTGGGCACC	97-112	119-134	-22
HLD39	rs17878444	1p22.1	CCTAAACAAAAATGGGAT	113-131	126-144	-13
HLD58	rs1610937	5q14.1	AGGA	78-82	137-141	-59
HLD67	rs1305056	5q33.2	CTACTGAC	77-86	140-148	-63
HLD84	rs3081400	8q24.12	CTTTC	113-118	137-142	-24
HLD88	rs8190570	9q22.32	CCACAAAGA	87-96	120-129	-33
HLD97	rs17238892	13q12.3	AGAGAAAGCTGAAG	79-93	96-110	-17
HLD99	rs2308163	14q23.1	TGAT	104-108	109-113	-5
HLD101	rs2307433	15q26.1	GTAG	77-81	131-135	-54
HLD111	rs1305047	17p11.2	CACA	109-113	122-126	-13
HLD124	rs6481	22q12.3	GTGGA	88-93	104-109	-16
HLD125	rs16388	22q11.23	ATTGCC	89-95	129-135	-40
HLD128	rs2307924	1q31.2	ATTAAATA	98-106	114-122	-16
HLD131	rs1611001	7q36.2	TTGGGCTTATT	81-94	102-113	-21
AMEL.X	M55418	Xp22.1-22.3	---	104	---	---
AMEL.Y	M55419	Yp11.2	---	110	---	---

Table 5.2 Table showing mini-INDEL primer sequences, along with their labelled fluorescence dyes and their working concentrations (* represents tailing of 5 bases to reverse primer of HLD67).

Marker (SNP #)	Primer Sequence	dye	Concentration (μM)
HLD6 (rs1610905)	Forward: GGAAGCGGTCTGGAAGTCAG Reverse: GGGTACCTCTGAGCTCATCC	FAM	0.5
HLD67 (rs1305056)	Forward: GAGATTTGGAGGACTGTGCATGT Reverse: GCTTCTTTCTGCCTCAGAACAAAAC**TGAAT***	FAM	0.5
HLD131 (rs1611001)	Forward: TTTTAGGCATTCTAATAGGACTTGTCT Reverse: AAACTTCTGTGAAGCTACTCAGTCT	FAM	0.5
AMELOGENIN	Forward: CCCTGGGCTCTGTAAAGAATAGTG-3' Reverse:ATCAGAGCTTAAACTGGGAAGCTG-3	FAM	0.5
HLD39 (rs17878444)	Forward: ACCTGACTTACTTGCCCAAAA Reverse: CCTTGGTAATTCAGCAACAATCT	VIC	0.5
HLD101 (rs2307433)	Forward: TTCTGCCCAATGTTCATCAG Reverse: CATGCATGGTGAGCAGGAG	VIC	0.5
HLD124 (rs6481)	Forward: AGGATTCCTGTGTTGGTTAGCA Reverse: TCACGAAGCCACAAGTATTTAC	VIC	0.5
HLD58 (rs1610937)	Forward: CATTTAGGAAGCCAAATAGGATG Reverse: AGTCTGCAGAAAATGGGTCA	NED	0.5
HLD88 (rs8190570)	Forward: GAGAAGCATCAGTGTTGGGGA Reverse: TGGGGCTCCACTTTAGTAGTC	NED	0.5
HLD84 (rs3081400)	Forward: CTGACTTTGTTGCTAGTTTGTCA Reverse: AGGAGCCCTGAATTATGTATCA	NED	0.5
HLD99 (rs2308163)	Forward: GTAGAGCTGGAGTTGAGAGTCG Reverse: TCAGGAATGGGTTTTGTTGTGG	NED	0.5
HLD97 (rs17238892)	Forward: CCTGGGCAACAAGAGTGAA Reverse: TCCGGCAGATAAGAAAATCAA	PET	0.5
HLD111 (rs1305047)	Forward: TGTTCACTGGCTAAACTATGTGTAT Reverse: AGCCCTCAAGTTAAGAATGATTT	PET	0.5
HLD125 (rs16388)	Forward: TGCCTCTTCGTCTCATCGAC Reverse: GGTTGGCTCTCGTTGGCAT	PET	0.5
HLD128 (rs2307924)	Forward: ATGGTTACCACCAAGAGTTACATT Reverse: CGCTAGCAGAGCTAATGTTTTGA	PET	0.5

5.2.7 PCR Primer Optimisation

All the INDELs were initially optimised as singleplex in order to evaluate single primer efficiency and expected amplicon size. In order to optimise singleplex reaction, Platinum multiplex PCR master mix (12.5 µl) from Life Technology was used for each reaction and following criteria were adopted for other factors;

(a) Primer concentration = 0.1 µM , 0.3 µM, 0.5 µM
(b) Annealing temperature = 55 °C , 58 °C, 60 °C
(c) Input amount of DNA (9947) = 0.3 ng , 0.5 ng , 1.0 ng

Table 5.3 Table representing three different thermal cycler programs used for the validation of mini- INDELs kit.

	Steps	Program A	Program B	Program C
Stage A	Denaturation	95 °C (2 min)	95 °C (2 min)	95 °C (2 min)
Stage B (30 cycles)	Denaturation Annealing Extension	95 °C (30 s) 55 °C (120 s) 72 °C (60 s)	95 °C (30 s) 58 °C (120 s) 72 °C (60 s)	95 °C (30 s) 60 °C (120 s) 72 °C (60 s)
Stage C	Final Extension	60 °C (60 min)	60 °C (60 min)	60 °C (60 min)
Stage D	Hold	4 °C (∞)	4 °C (∞)	4 °C (∞)

After optimisation as singleplex reaction, the amplification of the 14 mini-INDELs along with amelogenin was performed in a single multiplex PCR reaction by using following reaction mix of 25 µl;

Platinum master mix (life Technology) = 12.5 µl

Primer mix (0.5 µM each primer) = 2.5 µl

DNA (9947A) = 5.0 µl

Distilled water = 5.0 µl

The amplification reaction for the multiplex of 14 mini-INDELs and amelogenin was performed using a GeneAmp PCR system 9700 thermal cycler (Applied Biosystems™, UK). The thermal cycler 9700 was programmed using the conditions outlined in Table 5.4.

After the cycling protocol was completed, PCR products were stored at –20 °C.

Table 5.4 Table showing the validated PCR protocol for the mini-INDELs kit.

	Steps	Temperature	Time
Stage A	Denaturation	95 °C	2 min
Stage B (30 cycles)	Denaturation Annealing Extension	95 °C 58 °C 72 °C	30 s 120 s 60 s
Stage C	Final Extension	60 °C	60 min
Stage D	Hold	4 °C	∞

5.2.8 Capillary Electrophoresis

DNA fragment analysis was carried out on ABI 3130XL Prism Genetic Analysers using 36 cm capillaries and POP-4 polymer (Applied Biosystems™, UK). The parameters used during fragment analysis included, run temperature 60 °C, injection time 10 s, injection voltage 1.6 kVs, run voltage 15 kV.

For fragment analysis, samples were prepared by adding 1 µl of PCR product to 8.7 µl of Hi-Di formamide and 0.3 µl LIZ 500 size standard (Applied Biosystems™, UK). Thereafter, samples were denatured for 3 min at 95 °C and then placed on ice for 3 min.

5.2.9 Data Analysis

The data obtained from the capillary electrophoresis (CE) were analysed using GeneMapper Software v4.1 (Applied Biosystems™, UK). The panel and bin setting for mini-INDELs used in GeneMapper software are shown below in Tables 5.6 and 5.7. These panels and bins were developed to analyse the data for Genetic Analyser (GA) 3130XL. These panels and bins setting might vary to analyse data from other instruments like GA 310 and GA 3500.

Table 5.5 Table showing the Panel settings for mini-INDEL markers including dye colour, size (minimum and maximum) and ladder alleles.

Dye	Minimum Size	Maximum size	Ladder Alleles
Blue	77 bp	111 bp	D67-, D67+, D131-, D131+, D6-, X , Y , D6+,
Green	75 bp	130 bp	D101-, D101+, D124-, D124+, D122-, D122+, D39-, D39+,
Yellow	77 bp	116 bp	D58-, D58+, D88-, D88+, D99-, D99+, D84-, D84+
Red	77 bp	112 bp	D97-, D125-, D97+, D125+, D128-, D128+, D111-, D111+,

Table.5.6. Table indicating the Bins settings for mini-INDEL alleles for data generated from Genetic analyser 3130XL including their position, dyes and width.

Alleles	Bins position (bp)	Dyes	Bin Width (bp)	Alleles	Bins position (bp)	Dyes	Bin Width (bp)
D67-	70.5	Blue	0.5	**D58-**	77.5	Yellow	0.5
D67+	79.5		0.5	**D58+**	81.5		0.5
D131-	80.5		0.5	**D88-**	85.5		0.5
D131+	93.0		0.5	**D88+**	94.5		0.5
D6-	96.0		0.5	**D99-**	103.0		0.5
X	102.0		0.5	**D99+**	107.0		0.5
Y	108.0		0.5	**D84-**	111.0		0.5
D6+	110.0		0.5	**D84+**	116.0		0.5
D101-	75.5	Green	0.5	**D97-**	77.5	Red	0.5
D101+	79.5		0.5	**D125-**	87.5		0.5
D124-	86.5		0.5	**D97+**	92.0		0.5
D124+	91.5		0.5	**D125+**	93.5		0.5
D122-	100.5		0.5	**D128-**	96.5		0.5
D122+	110.5		0.5	**D128+**	104.5		0.5
D39-	111.5		0.5	**D111-**	108.0		0.5
D39+	129.5		0.5	**D111+**	112.0		0.5

5.3. RESULTS

5.3.1 Sequence Alignment

The Fasta sequences of 14 INDEL markers from different population groups (limited amount of sequence data available) were aligned using BioEdit Sequence Alignment Editor (Hall 2011). The results of sequence alignment for all of 14 INDEL markers indicated that flanking region around the point of insertion or/and deletion are same among different population groups. Figure 5.2 shows the sequence alignment for HLD 99 marker that clearly depicts the uniform alignment of 16 nucleotide sequences of flanking regions for HLD 99 marker.

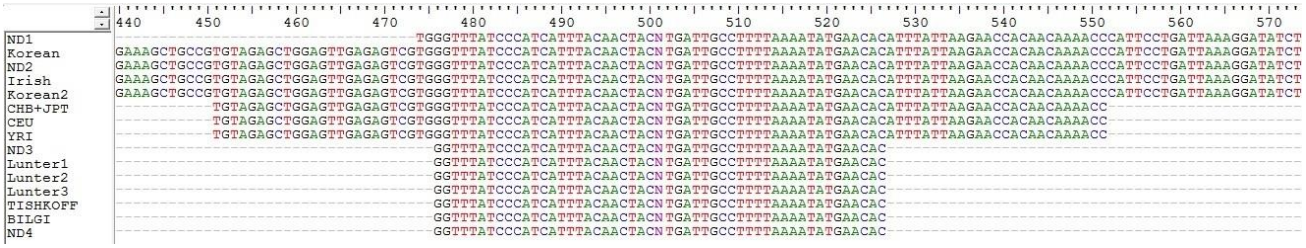

Figure 5.2 Figure showing the alignment of 16 sequences from different population groups for HLD 99 (rs2308163)

5.3.2 Singleplex Optimisation of mini-INDELs

All of the 14 mini-INDELS markers and the sex discriminating marker, amelogenin were optimised using control DNA 9947A at different PCR conditions (Table 5.3) in singleplex reactions.

5.3.2.1 Optimisation of Primer Concentration

Three different primer concentrations (0.1 µM, 0.3 µM, 0.5 µM) were tested to get the optimum concentration. By using 0.1 µM concentration of primer, the peak height was much lower than the threshold (100 RFU). The peak height for 0.3 µM was above the threshold but peak quality was not good, it was broader than the normal peak. The best peak height and quality were observed for 0.5 µM primer concentration (Figure 5.3).

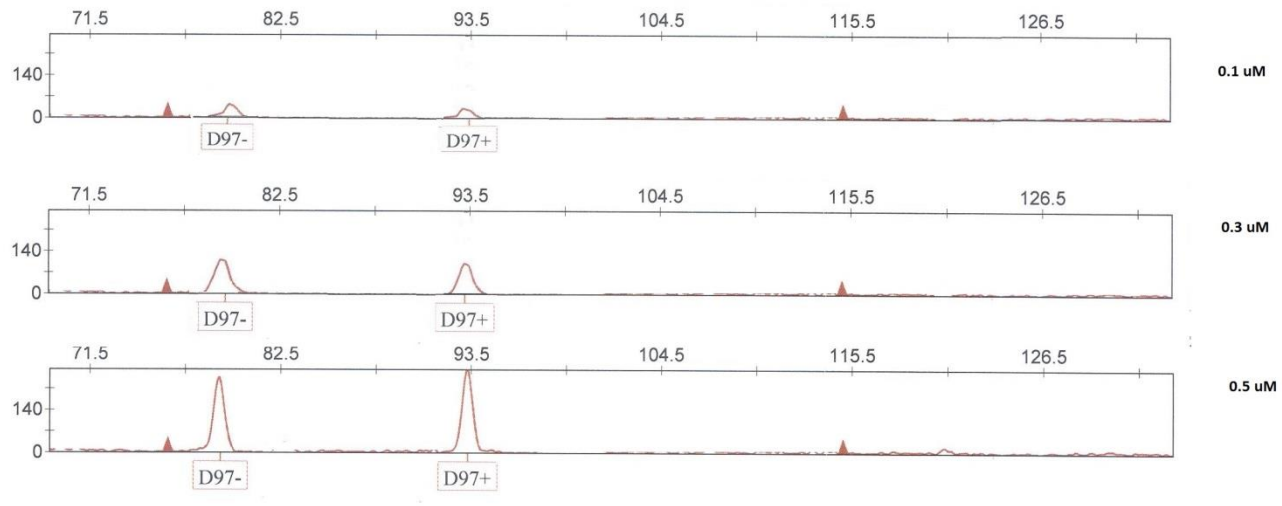

Figure 5.3 The original traces showing the optimization results of the amplication of D97 marker in singleplex form (n = 3) at three different primer concentrations (0.1 µM, 0.3 µM, 0.5 µM).

5.3.2.2 Optimisation of Annealing Temperature

Thermal cycler conditions were optimised in terms of annealing temperature (55 °C, 58 °C, and 60 °C). At 55 °C and 60 °C, some of the markers were either unable to amplify or have low amplification. The optimum amplification of all the INDEL markers occurred only at 58 °C.

5.3.2.3 Optimisation of Input DNA

In order to get the optimised quantity of DNA, three different amounts were used (0.3 ng, 0.5 ng, 1.0 ng). The amplification of most of markers was weak and drop-outs were observed while using 0.3 ng amount of input DNA. An input amount of 1 ng was too much for some of the markers, which resulted in over amplification in the form of split peaks and other artefacts. While 0.5 ng quantity of DNA gave the complete amplification of all the markers (Figure 5.4).

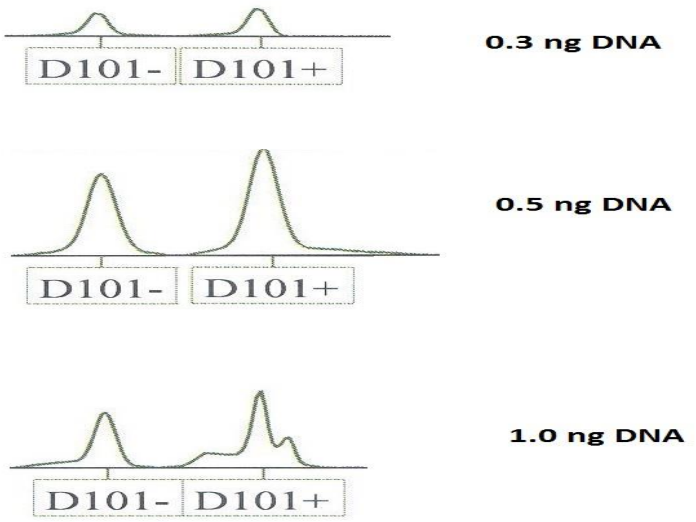

Figure 5.4 Original traces showing the optimization results of the amplication of D101 marker in singleplex form (n = 3) using three different amounts of DNA (0.3 ng, 0.5ng, 1.0ng).

In a summary, the best parameters were found to be the primer concentration of 0.5 µM, an annealing temperature of 58 °C and a quantity of DNA of 0.5 ng. The product size of all the markers was also matched with that of expected size while designing the primers. Electropherograms of the final results of individual INDEL marker as singleplex (14 mini-INDELS and amelogenin) are shown in Figures (5.5-5.8).

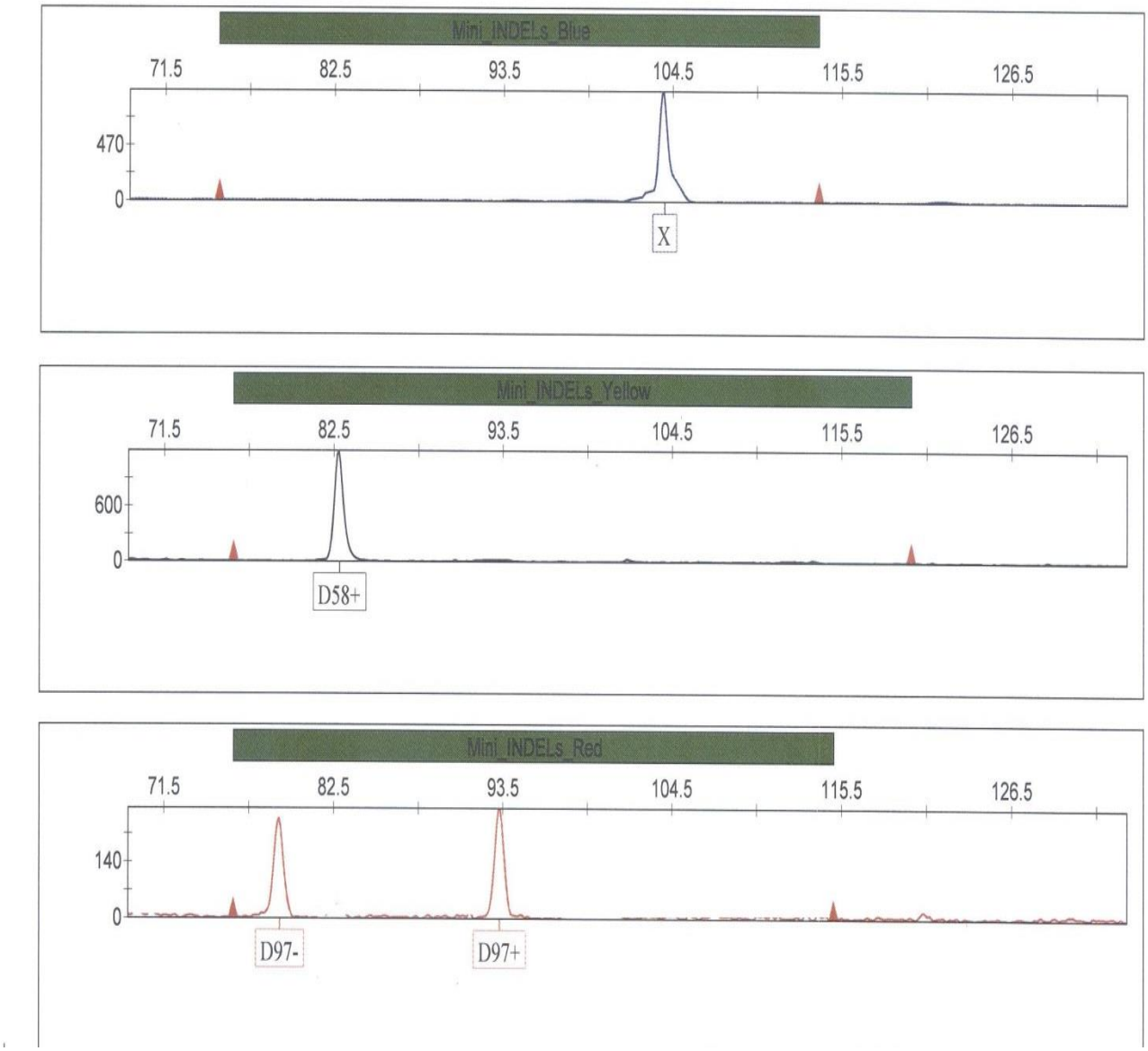

Figure 5.5 Original traces showing the electropherograms of the singleplex amplification results for the markers of amelogenin, HLD58 and HLD97 after final optimization with primer concentration of 0.5 µM, DNA quantity of 0.5 ng and an annealing temperature of 58 °C (n = 3).

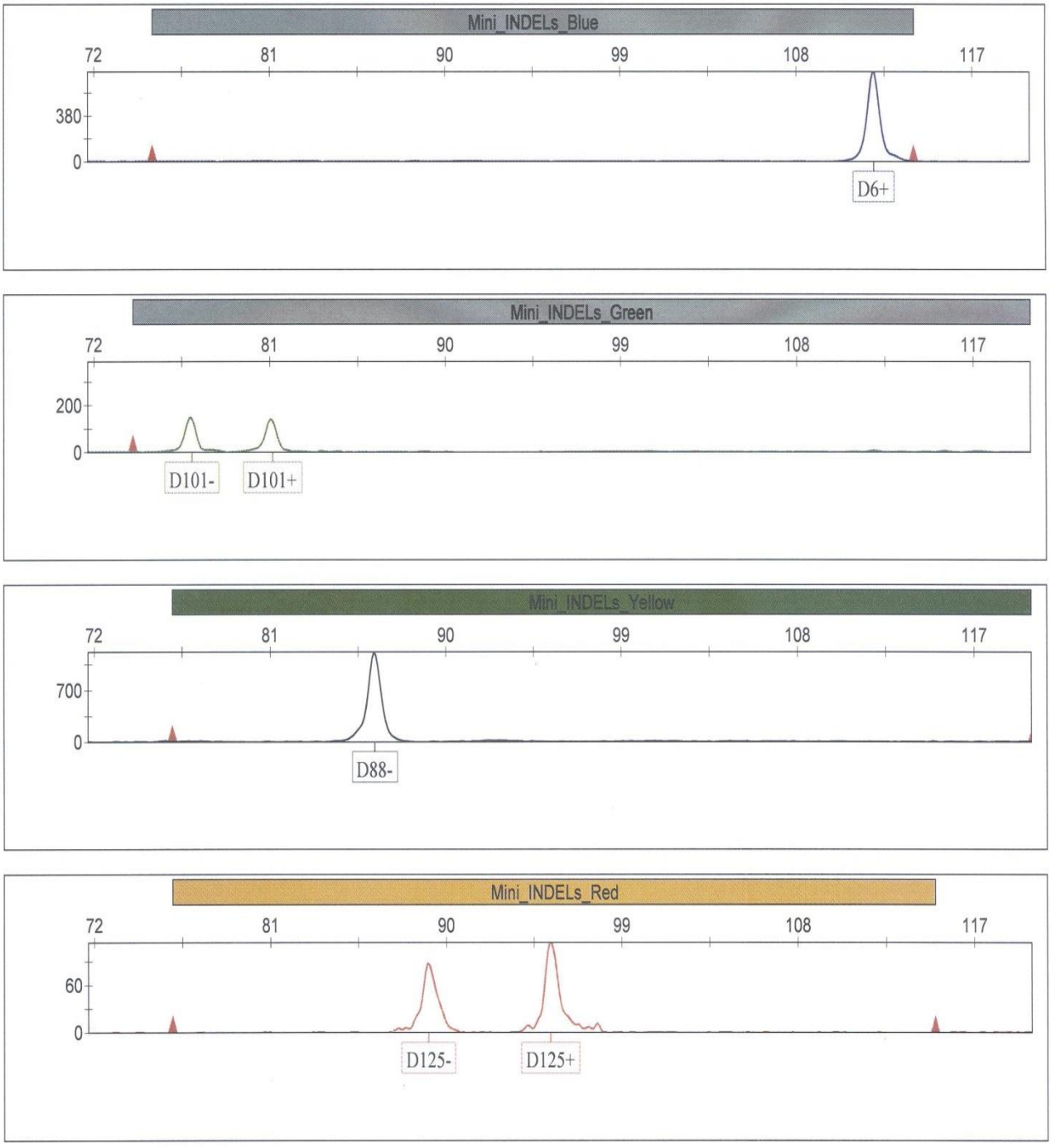

Figure 5.6 Original traces showing the electropherograms of singleplex amplification results for the markers of HLD6, HLD101, HLD88 and HLD125 after final optimization with primer concentration of 0.5 µM, DNA quantity of 0.5 ng and an annealing temperature of 58 °C (n = 3).

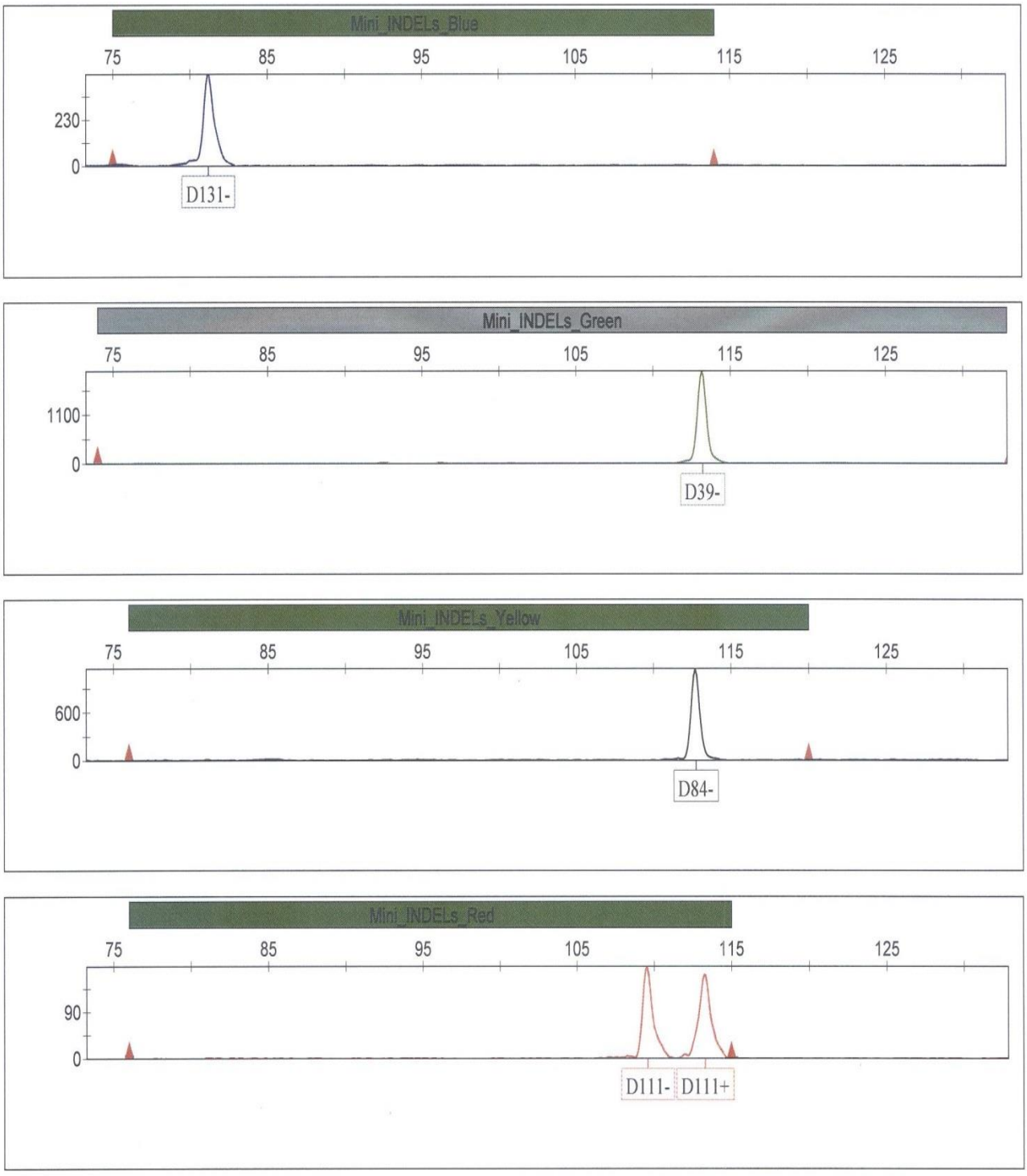

Figure 5.7 Original traces showing the electropherograms of singleplex amplification results for the markers of HLD131, HLD39, HLD84 and HLD111 after final optimization with primer concentration of 0.5 µM, DNA quantity of 0.5 ng and an annealing temperature of 58 °C (n = 3).

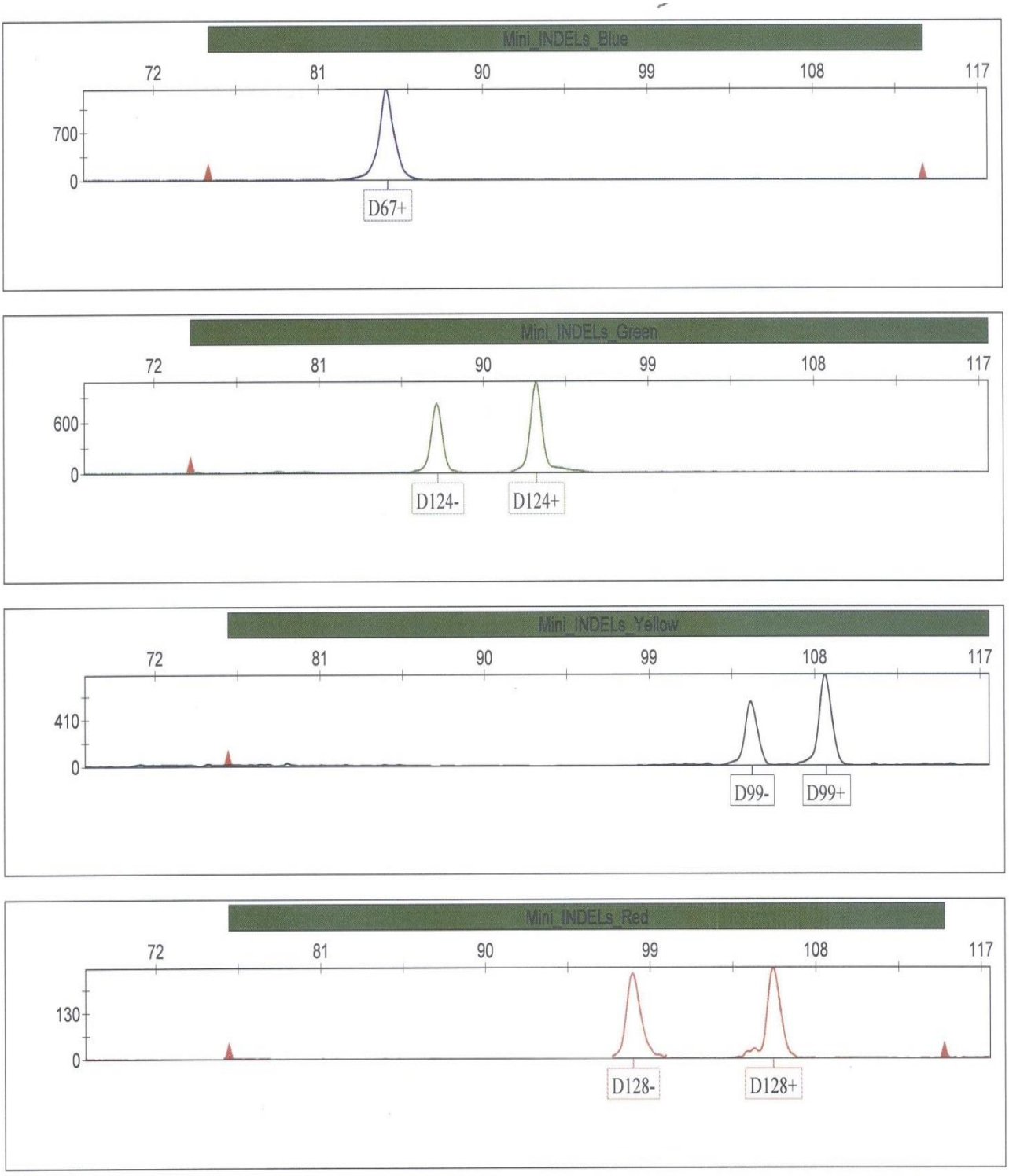

Figure 5.8 Original traces showing the electropherograms of singleplex amplification results for the markers of HLD67, HLD124, HLD99 and HLD128 after final optimization with primer concentration of 0.5 µM, DNA quantity of 0.5 ng and an annealing temperature of 58 °C (n = 3).

5.3.3 Development of the Multiplex

Following optimization as a singleplex, all the individual primers were mixed together in equal volumes to achieve a final 0.5 µM concentration for the primer mix. Then, a final amount of 2.5 µl of primer mix was used along with 12.5 µl of platinum reaction mix (Applied Biosystems™, UK) for a 25 µl PCR reaction. Subsequently, all the PCR reactions for concordance study and degradation/inhibition study were performed using same volume of reaction mix.

The control DNA 9947A (0.5 ng) was amplified by using the optimised PCR conditions of singleplex i.e. annealing temperature of 58 °C and final primer concentration of 0.5 µM (Figure 5.10). All the alleles of mini-INDEL markers for control DNA 9947A were fully concordant with those of DIPplex markers (see Chapter 4).

5.3.4. Redesigning of HLD67 reverse primer

During the analysis of some profiles generated from mini-INDELs kit, it was observed that alleles of HLD 131- and HLD 67+ were only one base pair apart which results in split peak pattern. To overcome this problem, a tail of 5 bases (TGAAT) was added to the reverse primer of HLD67. The primer was re-ordered and optimised again. After new primer was introduced into the multiplex, a distance of 3 base pair was observed between alleles of HLD 131- and HLD 67+ which resulted in two distinctive peaks (Figure 5.9).

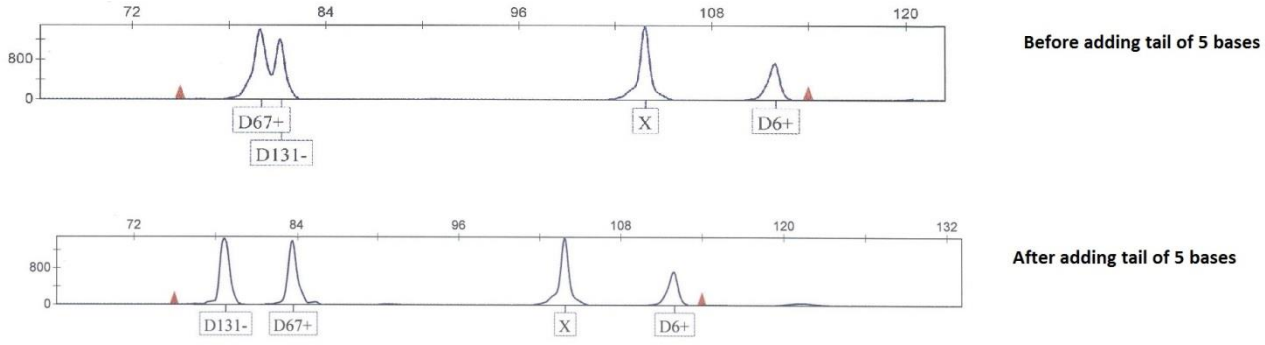

Figure 5.9 Original traces representing the re-arrangement of alleles of HLD67+ and HLD131- by adding a tail of 5 bases to reverse primer of HLD67 marker.

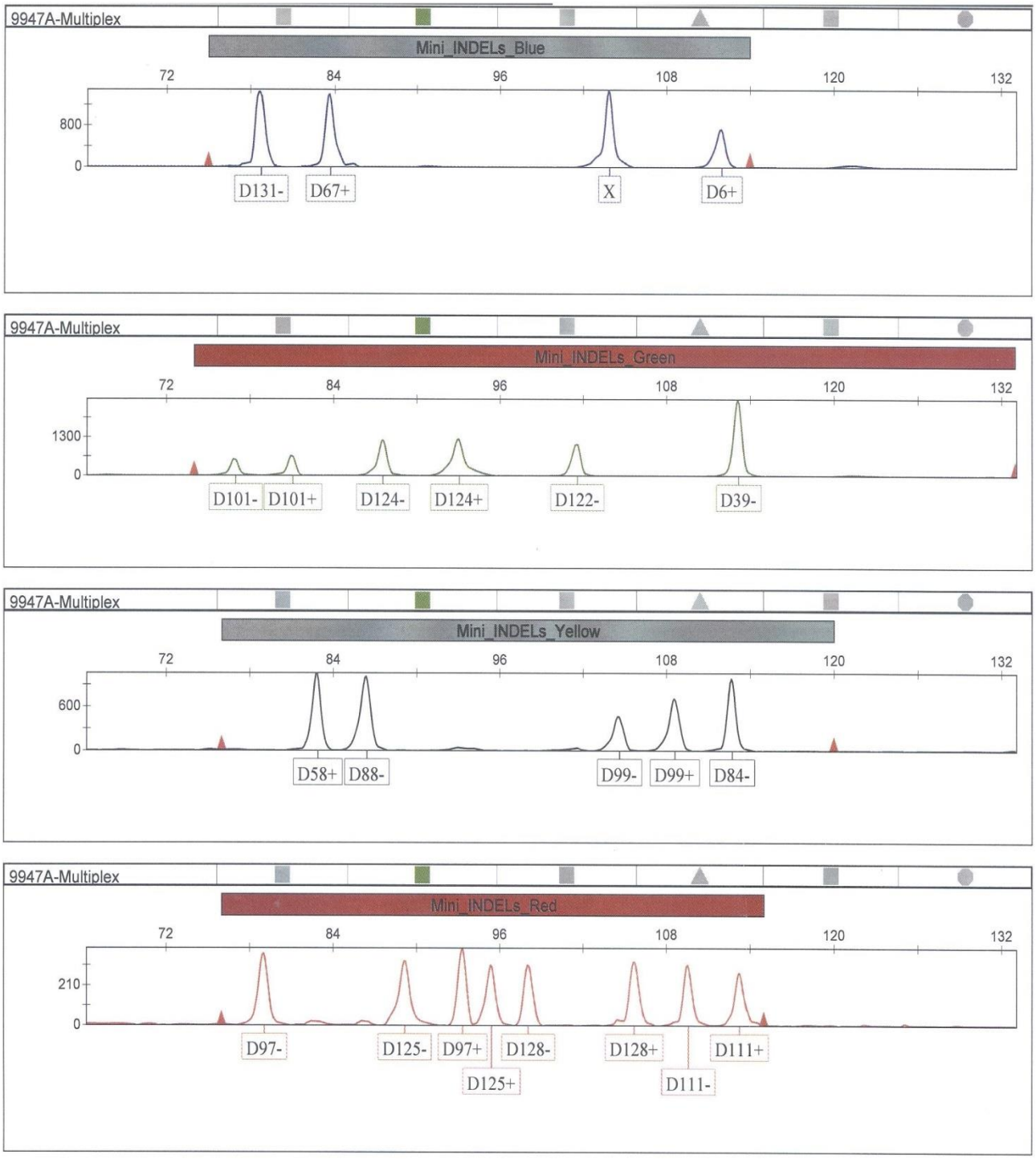

Figure 5.10 Original traces showing the electropherograms of amplification result of multiplex of control DNA9947A after final optimization with primer concentration of 0.5 µM, DNA quantity of 0.5 ng and an annealing temperature of 58 °C (n = 3).

5.3.5. Development of allelic ladder

An allelic ladder can be defined as an artificial mixture of available alleles present for a particular marker (Sajantila *et al.* 1992). Using same primers as used for other DNA samples can generate the alleles for the ladders. The allelic ladder provides a reference or standard of all alleles for each locus (Smith *et al.* 1995).

Each marker of mini-INDELs contains a maximum of two alleles in a combination of either allele of insertion or deletion (homozygous) or both together (heterozygous). The construction of allelic ladder is quite simpler for mini-INDELs due to presence of two types of alleles (insertion as + and deletion as -) as compared to STRs, which have multiple alleles (Sajantila *et al.* 1992; Smith *et al.* 1995).

In order to construct the allelic ladder for mini-INDELs, those DNA samples must be selected which on combining contained all of heterozygous alleles. Then, their quantification must be normalized to get the balanced alleles and later on amplified together in a single PCR reaction. To produce the additional quantities of allelic ladder of mini-INDELs, its amount must be diluted 1000 times with deionized water and was re-amplified (Baechtel *et al.* 1993). The amplified product could be detected using capillary electrophoresis with Genetic Analyser 3130XL. The data could be analysed using Gene Mapper® ID-X (Figure 5.8).

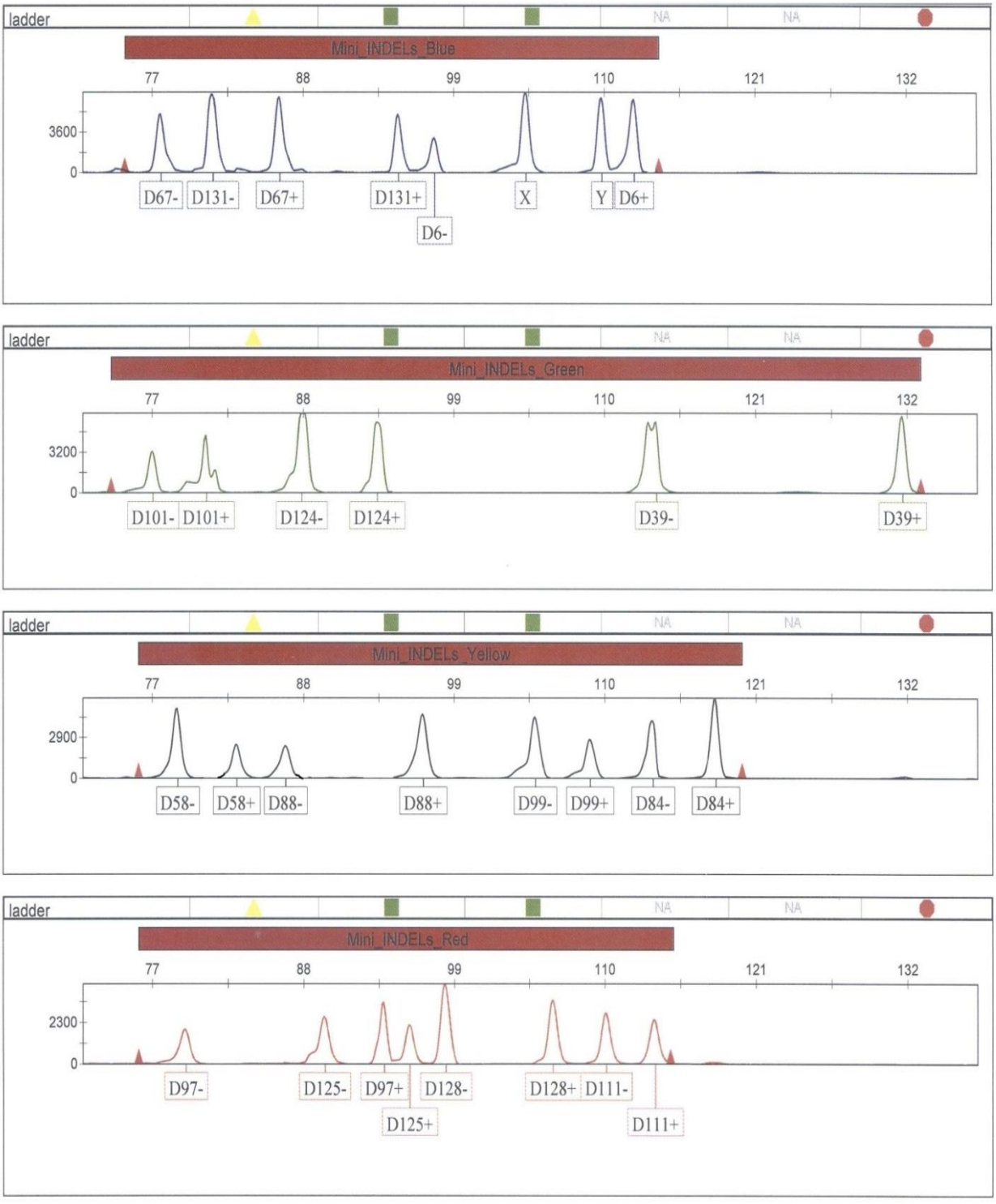

Figure 5.11 Original electropherogram showing all of 30 alleles present in the ladder of mini-INDELs kit (n = 3).

5.3.6. Sensitivity Study

A sensitivity study was designed to determine the minimum amount of DNA at which samples could generate a good quality profile with sufficient peak height ratio and well-balanced peaks. The sensitivity study was performed using Genetic analyser 3500.

DNA samples (9947A) was serially diluted upto amounts of 1 ng, 0.5 ng, 0.25 ng, 0.125 ng, 0.0625 ng, 0.03125 ng and 0.0156 ng in the replicates of three and they were subjected to amplification with Applied Biosystems™ Identifiler Plus®, Applied Biosystems™ Y-Filer®, Applied Biosystems™ MiniFiler® and mini- INDELs kits.

5.3.6.1 Sensitivity Study of Identifiler Plus® Kit

The sensitivity results from Identifiler plus® indicated that its loci were capable of amplifying efficiently upto 0.125 ng with good peak heights without any drop-out (Table.5.7). The drop-outs were observed in the DNA sample containing 0.0625 ng (4 alleles drop-out), 0.03125 ng (10 alleles drop-out) and 0.0156 ng (18 alleles drop-out). Figure 5.12 shows the profiles generated from Identifiler Plus® kit for its sensitivity at different amounts of DNA.

Table 5.7 Table showing the sensitivity results of Identifiler Plus® kit for different quantity of input DNA including the number of drop-outs from total number of alleles and average peak heights (n =3).

DNA concentration (ng)	# of alleles drop-out (out of 29 alleles)	Average Peak Heights (RFU)	STD. Deviation Peak Heights (RFU)
0.0156	18	75	20
0.03125	10	150	50
0.0625	4	300	100
0.125	0	700	150
0.250	0	1000	200
0.500	0	1500	350
1.00	0	3000	600

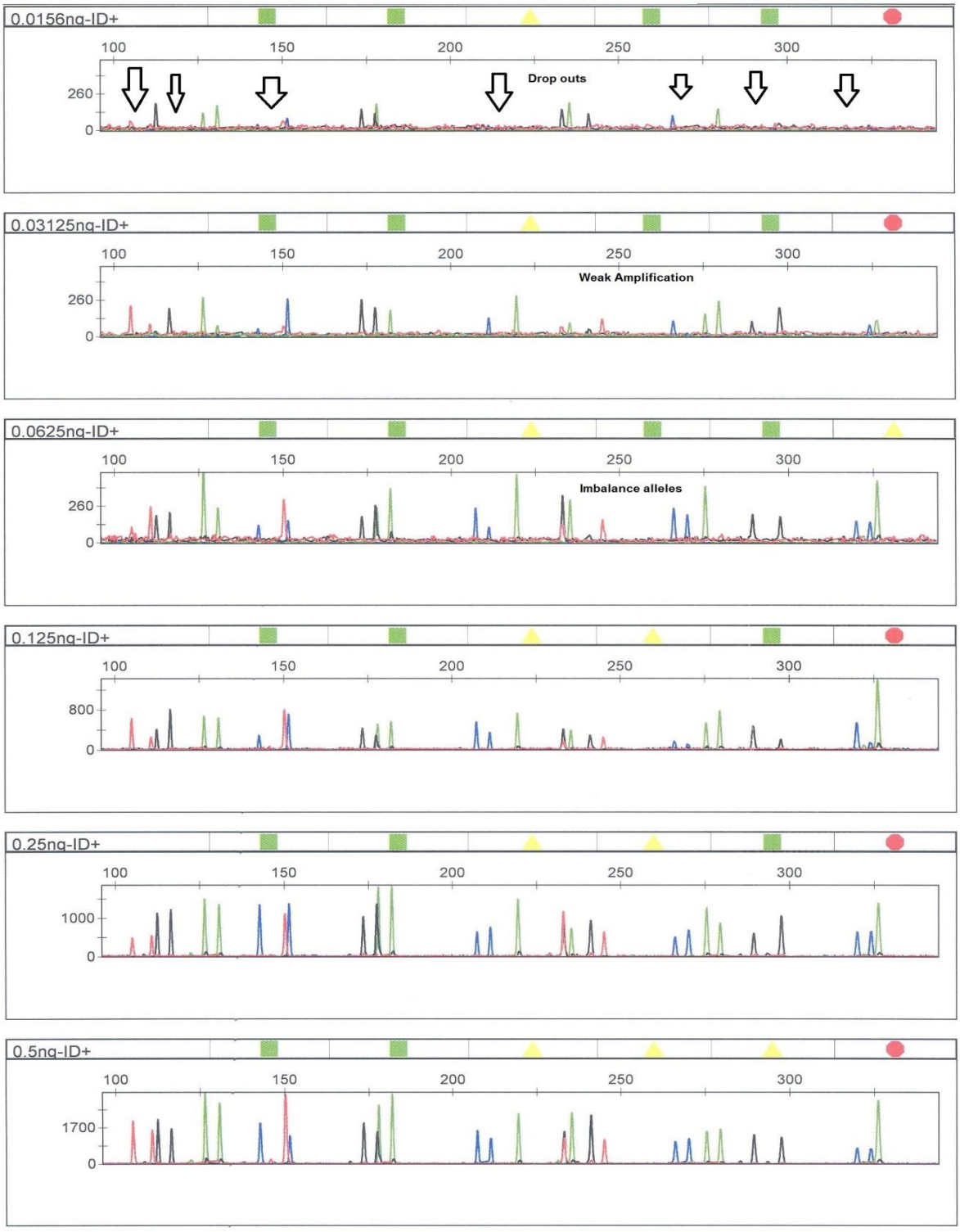

Figure 5.12 Original electropherogram representing the sensitivities results (n = 3) of Identifiler Plus® kit using 27 cycles recommended by manufacturer for different amount of DNA (0.0156 ng- 0.5 ng).

5.3.6.2 Sensitivity Study of Y-Filer® Kit

In this study, the results indicated that Y-Filer® kit shows a different pattern in its sensitivity. It showed well-balanced peaks at 0.25 ng and 0.125 ng without any drop outs. While at higher amounts of 1 ng and 0.5 ng and at lesser amounts of 0.625 ng and 0.03125 ng, the results indicated significant number of drop-outs (Table 5.8). In contrast, the amount of 0.0156 ng showed a weak profile with average peak height of 70 RFU and 10 drop outs (Figure 5.13).

Table 5.8 Table showing the sensitivity results of Y-Filer® kit for different quantity of input DNA including the number of drop-outs from total number of alleles and average peak heights (n=3).

DNA concentration (ng)	# of alleles drop-out (out of 17 alleles)	Average Peak Heights(RFU)	STD. Deviation Peak Heights (RFU)
0.0156	10	70	20
0.03125	5	300	50
0.0625	4	600	100
0.125	0	1000	150
0.250	0	2000	210
0.500	7	2500	370
1.00	11	3500	620

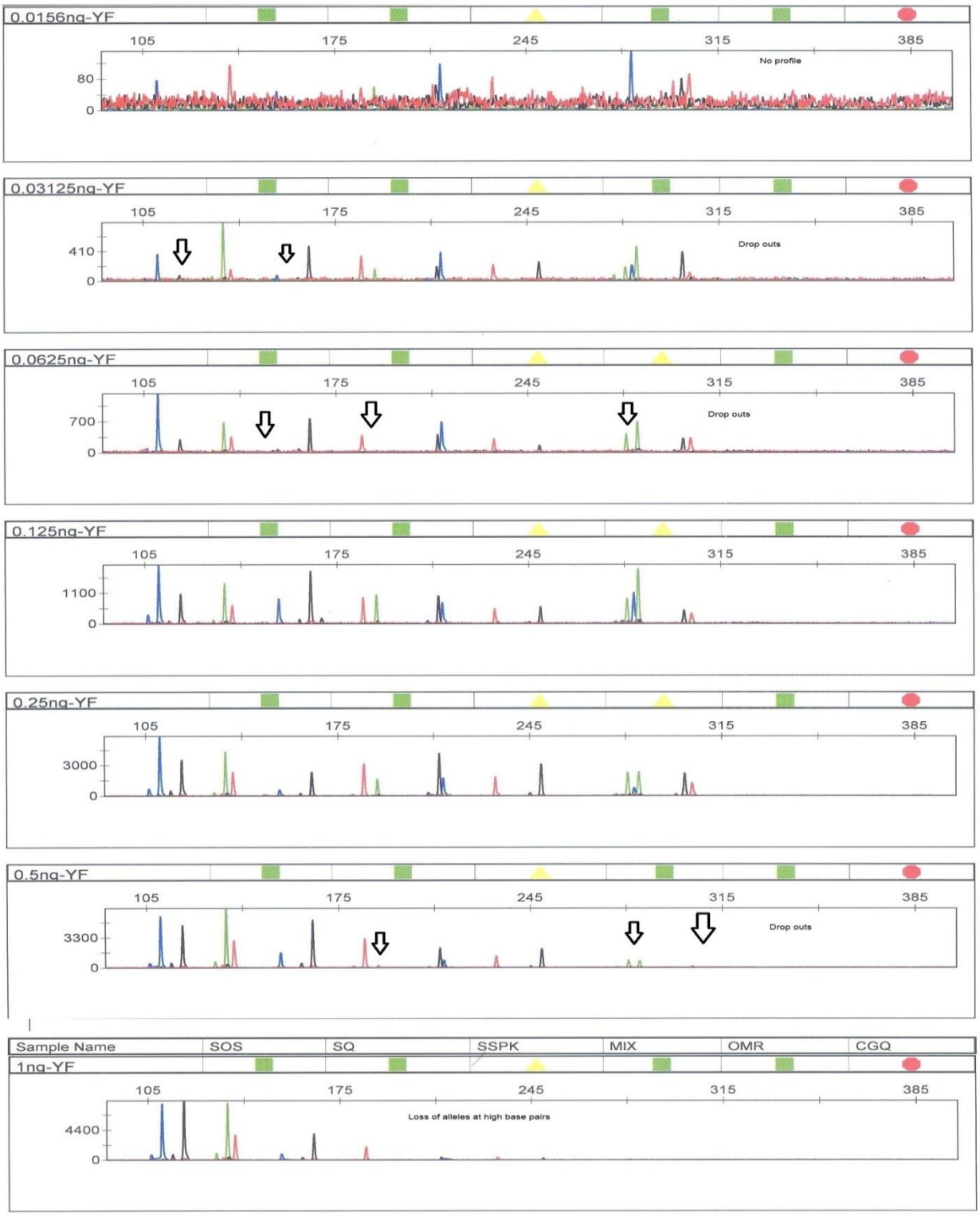

Figure.5.13. Original electropherogram representing the sensitivities results of Y-Filer kit (n = 3) using 30 cycles recommended by manufacturer for different amount of DNA (0.0156 ng-1.0 ng).

5.3.6.3 Sensitivity Study of MiniFiler® Kit

In this study, the results showed that MiniFiler® kit was more sensitive than that of Identifiler Plus® kit as it produces full profiles with good peak heights and without any drop-outs upto 0.0625 ng. However, the data showed weak amplification of some loci but without any drop-out for 0.03125 ng. The sample containing 0.0156 ng shows 2 drop-outs of alleles but with good peak height (Table 5.9 and Figure 5.14).

Table 5.9 Table showing the sensitivity results of MiniFiler® kit for different quantity of input DNA including the number of drop-outs from total number of alleles and average peak heights (n = 3).

DNA concentration (ng)	# of alleles drop-out (out of 16 alleles)	Average Peak Heights(RFU)	STD. Deviation Peak Heights (RFU)
0.0156	2	400	90
0.03125	0	800	150
0.0625	0	1500	400
0.125	0	2500	550
0.250	0	5000	700
0.500	0	8000	850
1.00	0	12000	950

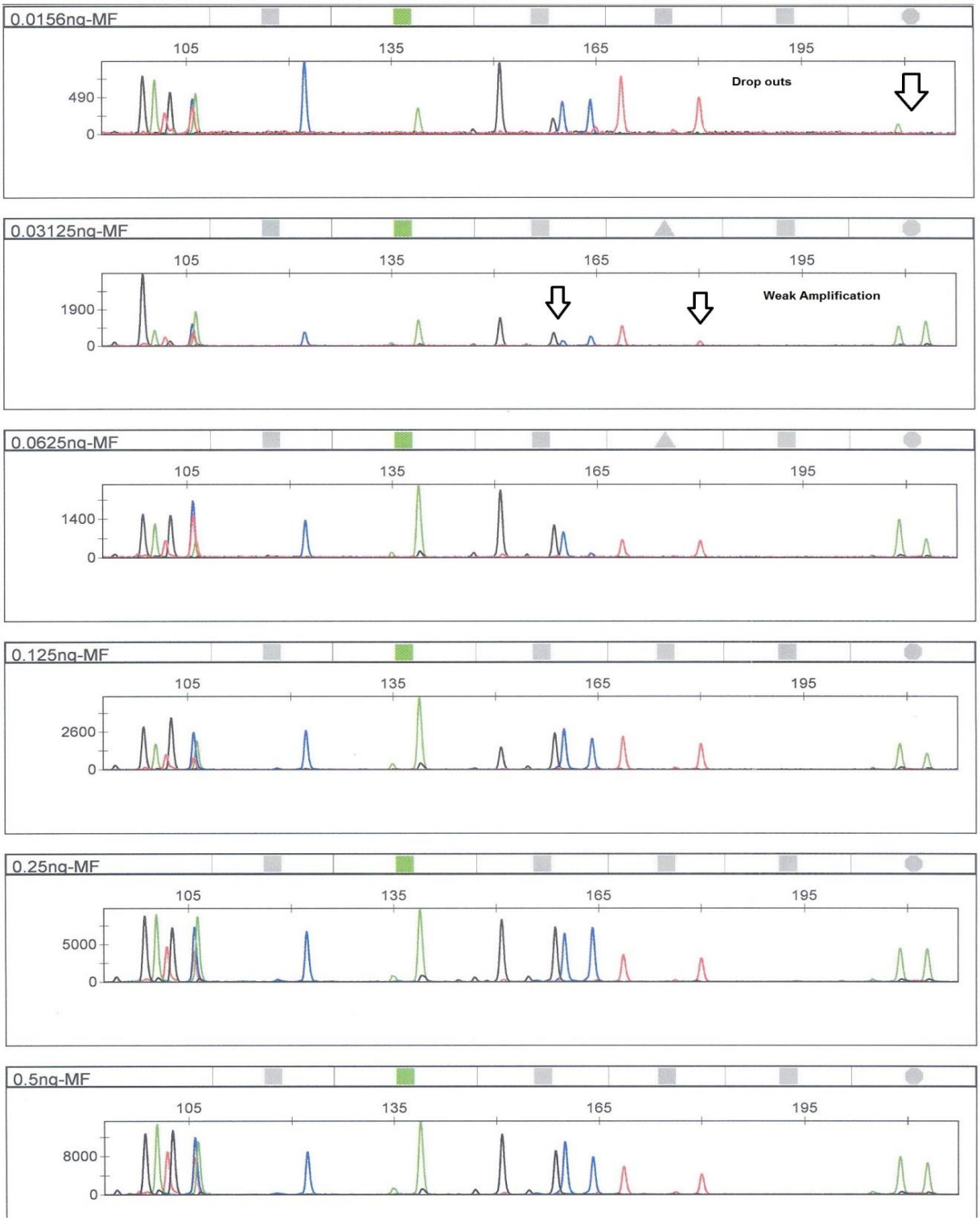

Figure 5.14 Original electropherogram indicating the sensitivities results of MiniFiler® kit (n = 3) using 30 cycles recommended by manufacturer for different amount of DNA (0.0156 ng-0.5 ng).

5.3.6.4 Sensitivity Study of mini-INDELs Kit

The sensitivity results of mini-INDELs kit were found to be better than those of all the three Life Technology kits. It produces full and balanced profiles for all the samples with good peak heights and without any drop-outs. While weak amplification of some loci was observed only for sample with 0.0156 ng concentration (Table 5.10 and Figure 5.15).

Table 5.10 Table showing the sensitivity results of mini-INDELs kit for different quantity of input DNA including the number of drop-outs from total number of alleles and average peak heights (n = 3).

DNA concentration (ng)	# of alleles drop-out (out of 20 alleles)	Average Peak Heights(RFU)	STD. Deviation Peak Heights (RFU)
0.0156	0	300	40
0.03125	0	500	80
0.0625	0	900	190
0.125	0	1200	250
0.250	0	3500	470
0.500	0	4500	600
1.00	0	11000	1050

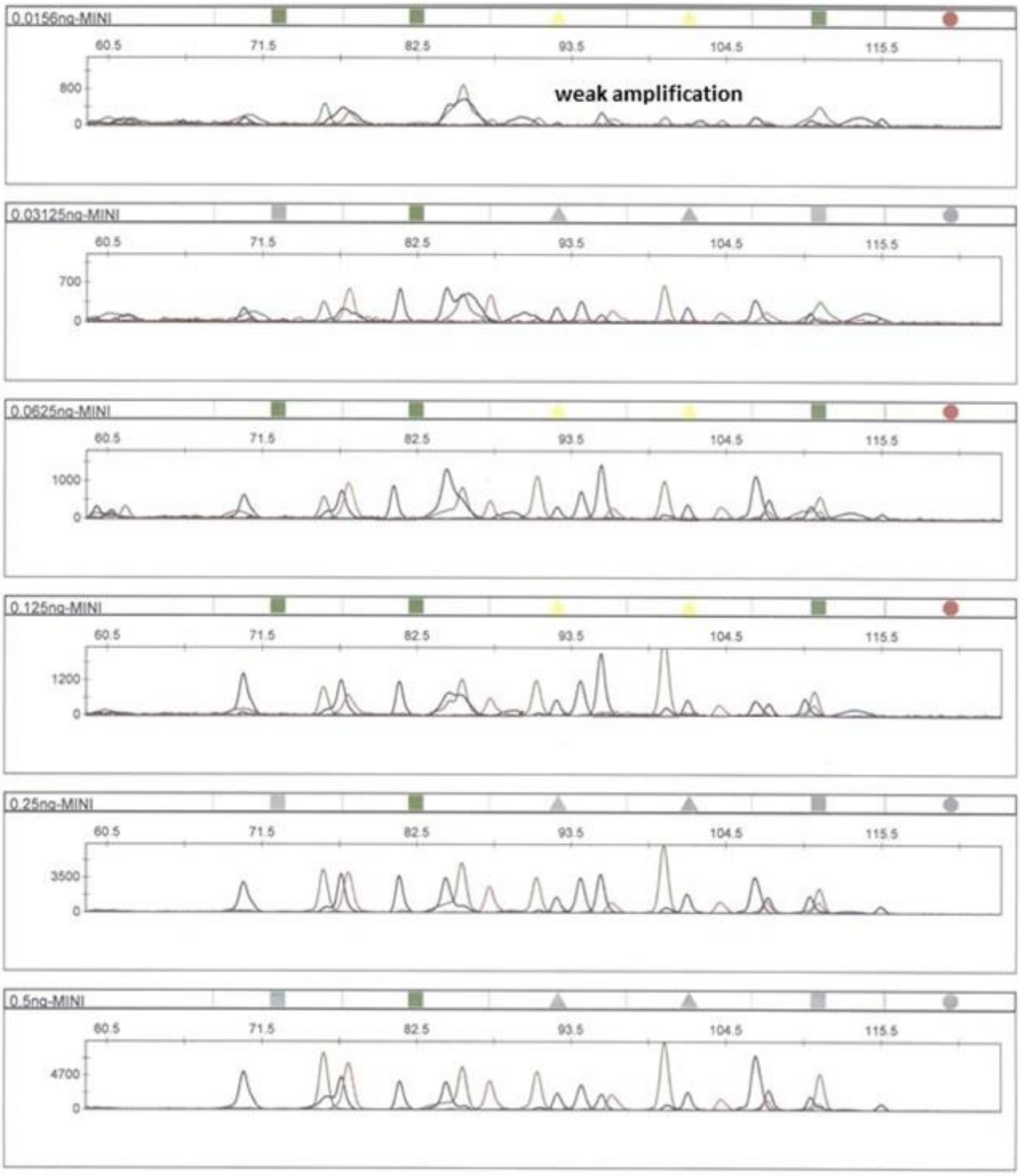

Figure 5.15 Original electropherogram representing the sensitivities results of mini-INDELs kit (n = 3) using 30 cycles for different amount of DNA (0.0156 ng- 0.5 ng).

5.3.6.5 Sensitivity Study of Investigator® DIPplex Kit

The sensitivity test for Qiagen™ Investigator® DIPplex kit was earlier mentioned in Chapter 3. The results have indicated that its loci were capable of amplifying efficiently for 0.5 ng with good peak heights and quality. But drop outs of 2 alleles were also observed for each of the samples 0.3 ng and 0.1 ng (Figure 5.16).

Table 5.11 Table showing the sensitivity results of Investigator® DIPplex kit for different quantity of input DNA including the number of drop outs from total number of alleles and average peak heights (n = 3).

DNA concentration (ng)	# of alleles drop out (out of 13 alleles)	Average Peak Heights(RFU)	Std. Deviation Peak Heights(RFU)
0.1	2	900	230
0.3	2	1500	550
0.5	0	2500	900

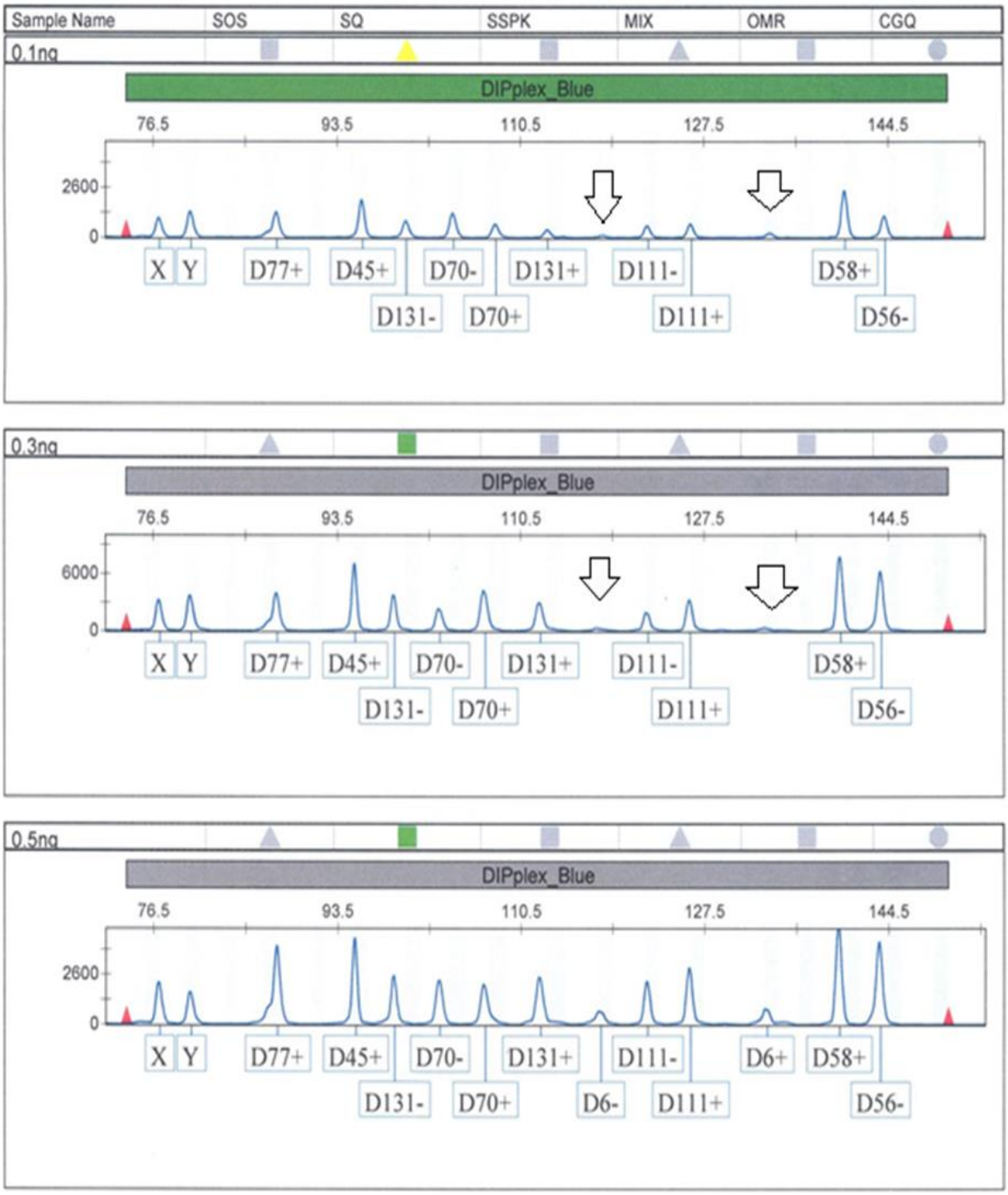

Figure 5.16 Original electropherogram represents the sensitivities results of Investigator® DIPplex kit (n = 3) using 30 cycles for different amount of DNA (0.01ng-0.5 ng). The figure is adopted from Chapter 3.

5.3.6. Concordance Study

A total of 175 DNA samples from four different populations (Qatari, Sudanese, Tunisian and Yemeni), previously analysed with Qiagen™ DIPplex kit were processed with mini-INDELs kit to assess the concordance. The results showed that the concordance between Qiagen™ Investigator® DIPplex kit and mini-INDELs kit was observed in 99.7% (4365 out of 4375) of INDEL allele calls. The 10 differences are listed in table 5.12 and they encompass the loci HLD 39 (n = 1), HLD58 (n = 1), HLD 67 (n = 3), HLD84 (n = 1), HLD97 (n = 1), HLD125 (n = 1) and HLD131 (n = 2). All other 7 loci and the amelogenin were fully concordant for all the studied samples. Four original electropherograms are shown below the concordance and non-concordance between both kits (Figures 5.17-5.20).

Table 5.12 Table showing the details of non-concordant alleles (n = 10) for Investigator® DIPplex and mini-INDELs kits.

Population groups	Sample ID	Mini INDELs Alleles	DIPplex Alleles
Qatari	QAT-001	D125-	Null Allele
	QAT-024	D97-	Null Allele
	QAT-093	D131+	Null Allele
	QAT-095	D39+	Null Allele
	QAT-098	D67+	Null Allele
Tunisian	TUN-018	Null Allele	D84-
	TUN-035	D131+	Null Allele
	TUN-096	D67+	Null Allele
	TUN-097	D67+	Null Allele
Yemeni	YEM-090	Null Allele	D58+
Sudanese	-	-	-
		10 non-concordant alleles	

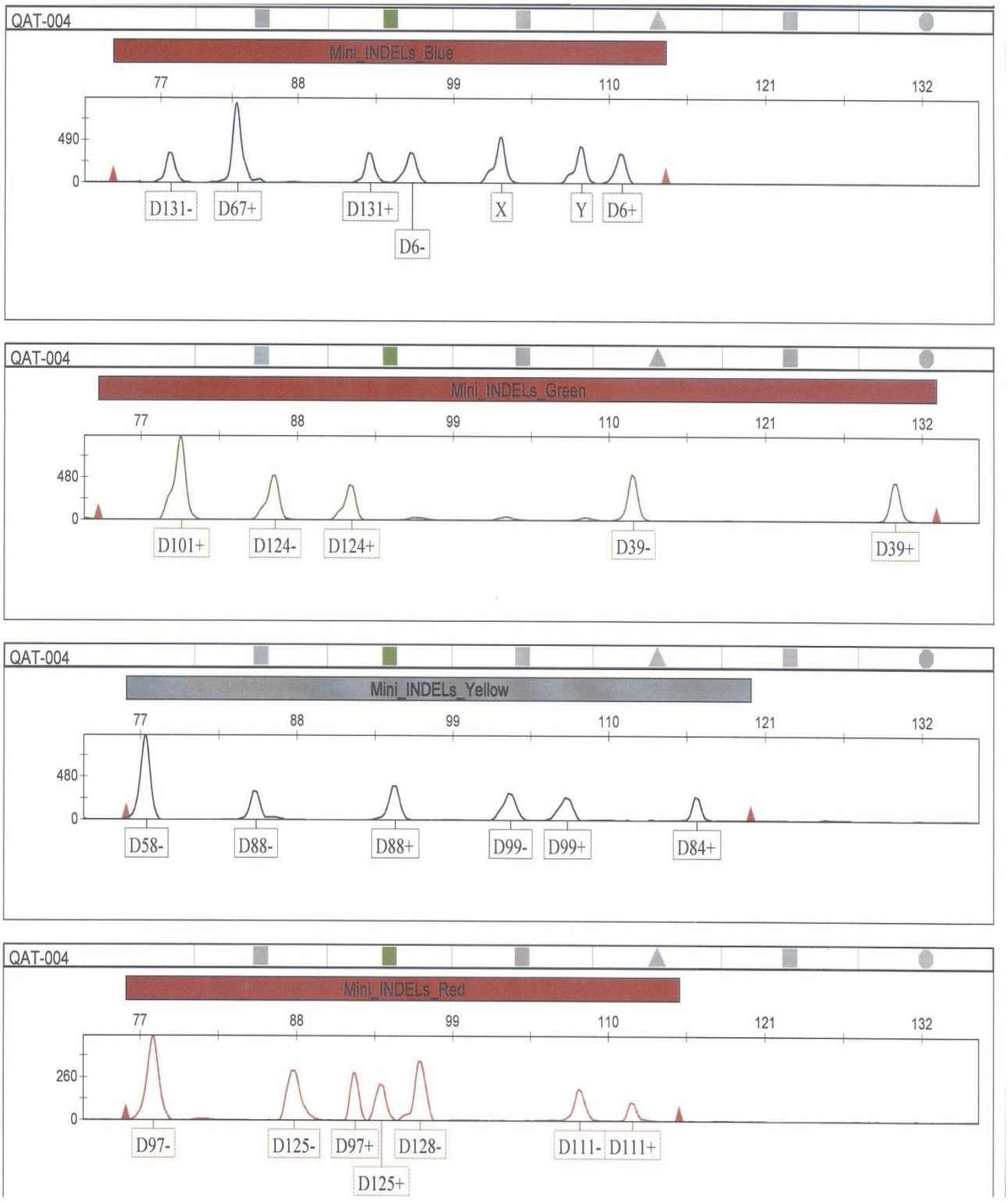

Figure 5.17 Original electropherogram showing the profile of a Qatari DNA sample (QAT-004) using mini-INDELs kit (n = 50) to show the concordance with that of DIPplex kit.

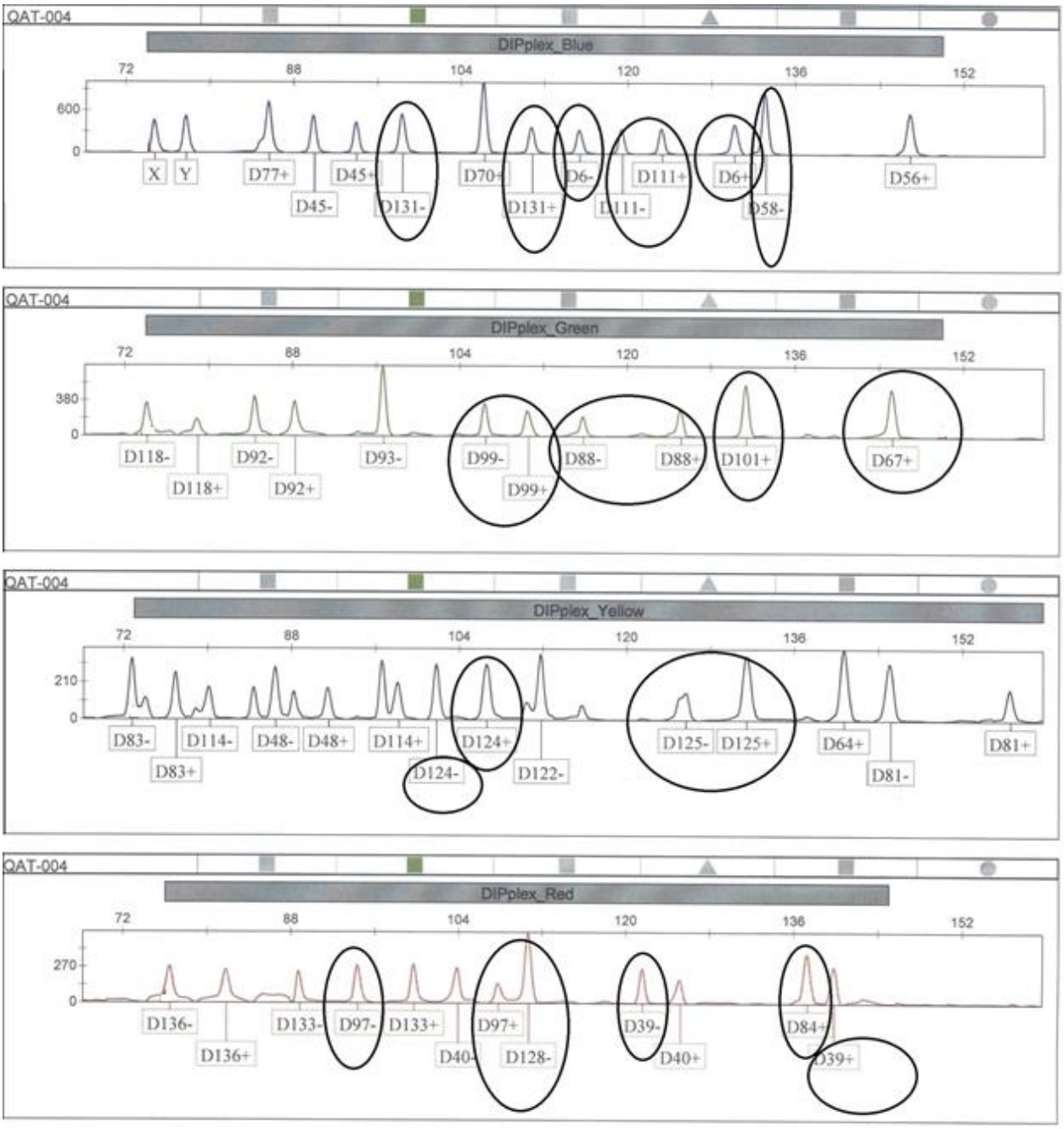

Figure 5.18 Original electropherogram showing the profile of a Qatari DNA sample (QAT-004) using Investigator® DIPplex kit (n = 100) representing 100% concordance of all alleles with those of mini-INDELs kit (circles representing the alleles which are concordant in both kits).

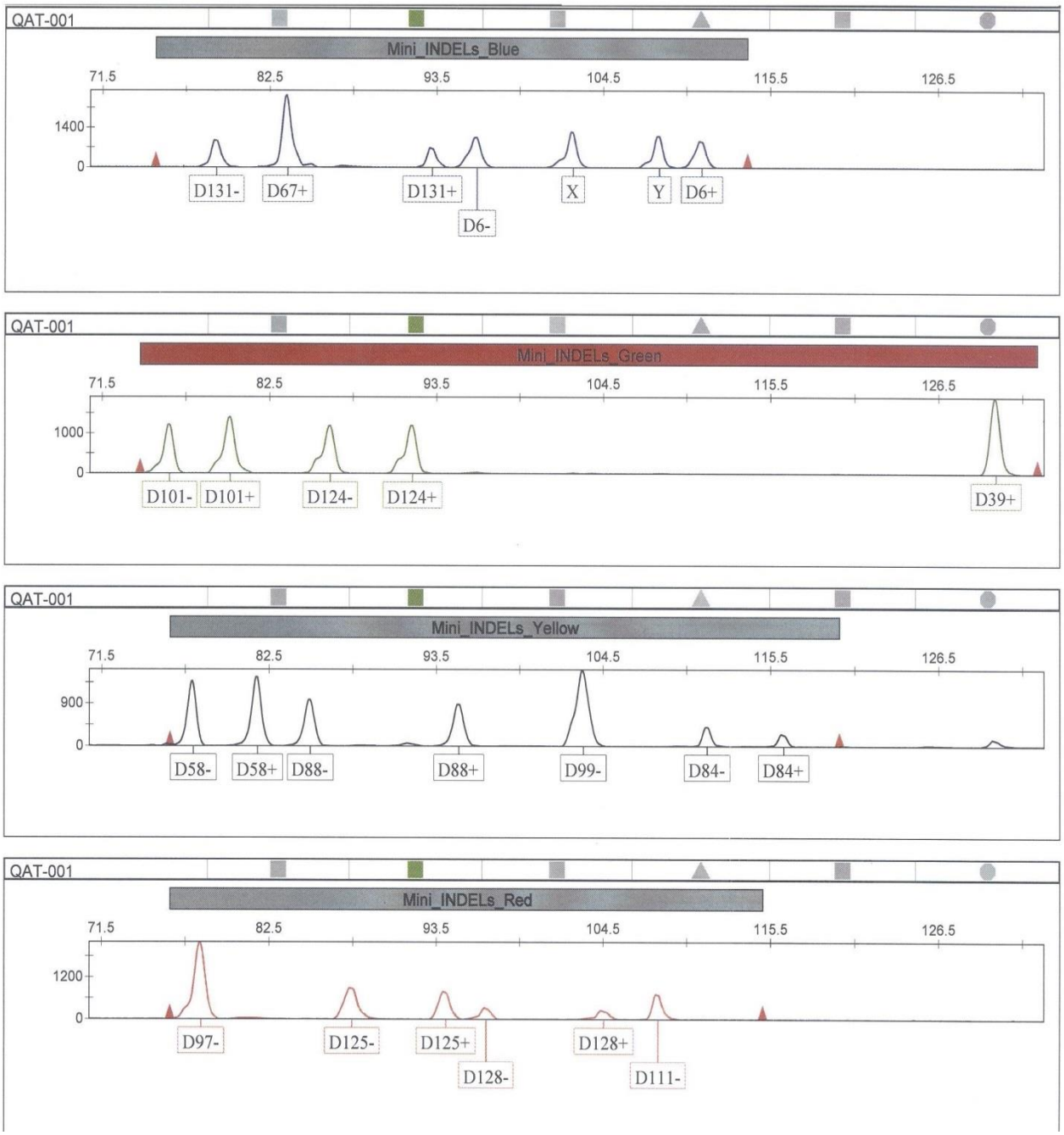

Figure 5.19 Original electropherogram showing the profile of a Qatari DNA sample (QAT-001) using mini-INDELs kit (n = 50) to show the non-concordance of D125- allele with that of Investigator® DIPplex kit.

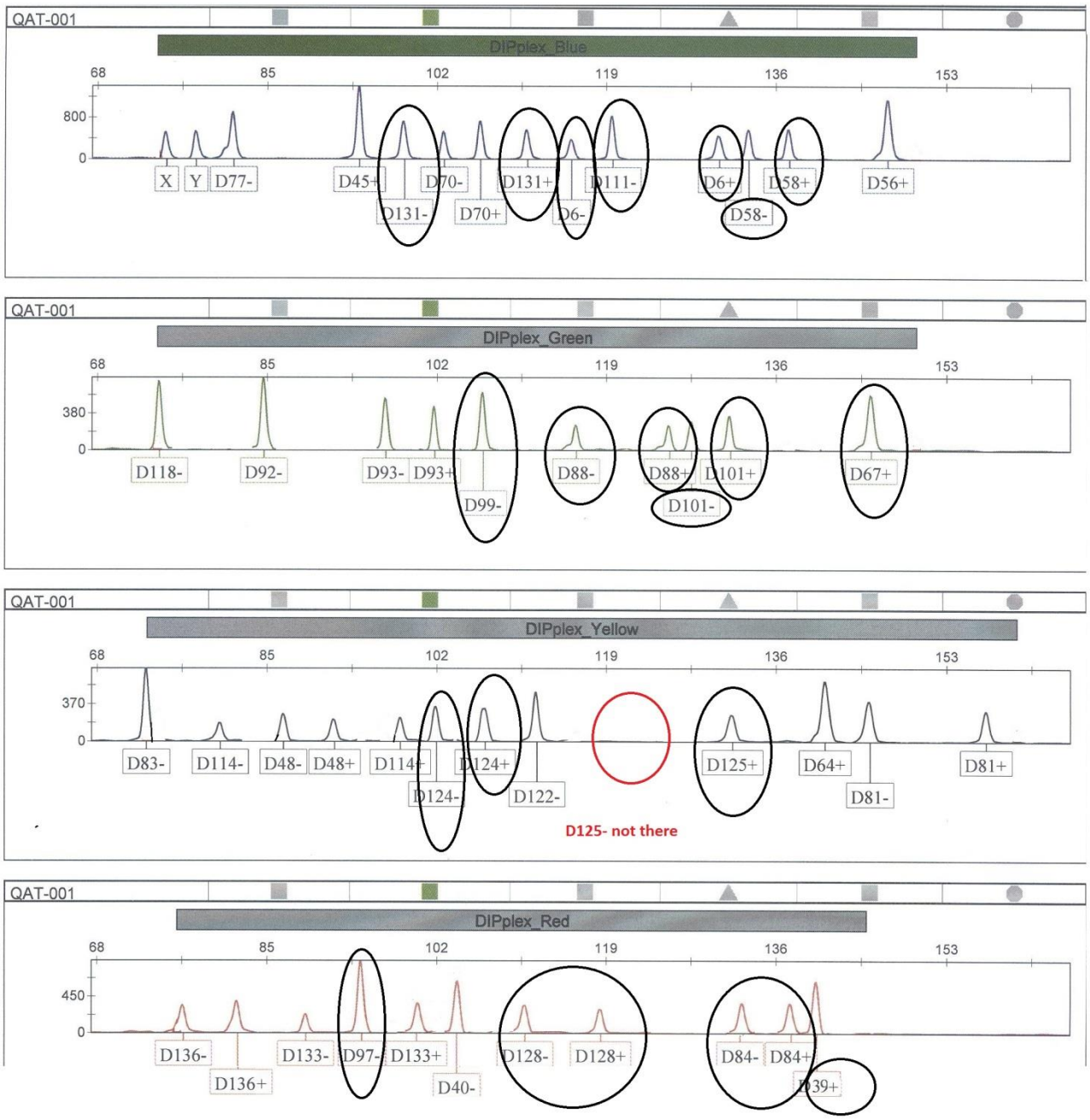

Figure 5.20 Original electropherogram showing the profile of Qatari DNA sample (QAT-001) using Investigator® DIPplex kit (n = 100) representing non-concordance of one allele (D125-) with that of mini-INDELs kit (circles representing the alleles which are concordant in both kits except D125-).

5.3.7. Comparison Study with DIPplex kit

A comparison study was designed to compare the efficiency of Qiagen™ Investigator® DIPplex kit and mini-INDELs kit. A total of 25 DNA samples were selected which were previously analysed with DIPplex kit with no or little amplification and were subjected to amplification with mini-INDELs kit.

The results showed that mini-INDELs kit was capable of amplifying all of 25 samples efficiently. The results of this study clearly have indicated that mini-INDELs kit is more robust and sensitive than Investigator® DIPplex kit.

The profiles from both kits were also tested for peak height ratios. The mean peak height ratios and quality of peaks of the profiles generated from mini-INDELs kit were found to be much better than those of Investigator® DIPplex kit.

Figures 5.21 and 5.22 show the two profiles of the same sample (QAT-008). The profile generated using Investigator® DIPplex kit showed no amplification of any of the loci except some artefacts. On the other hand, when the same sample with same amount of DNA was amplified with mini-INDELs, the data produced the complete profile.

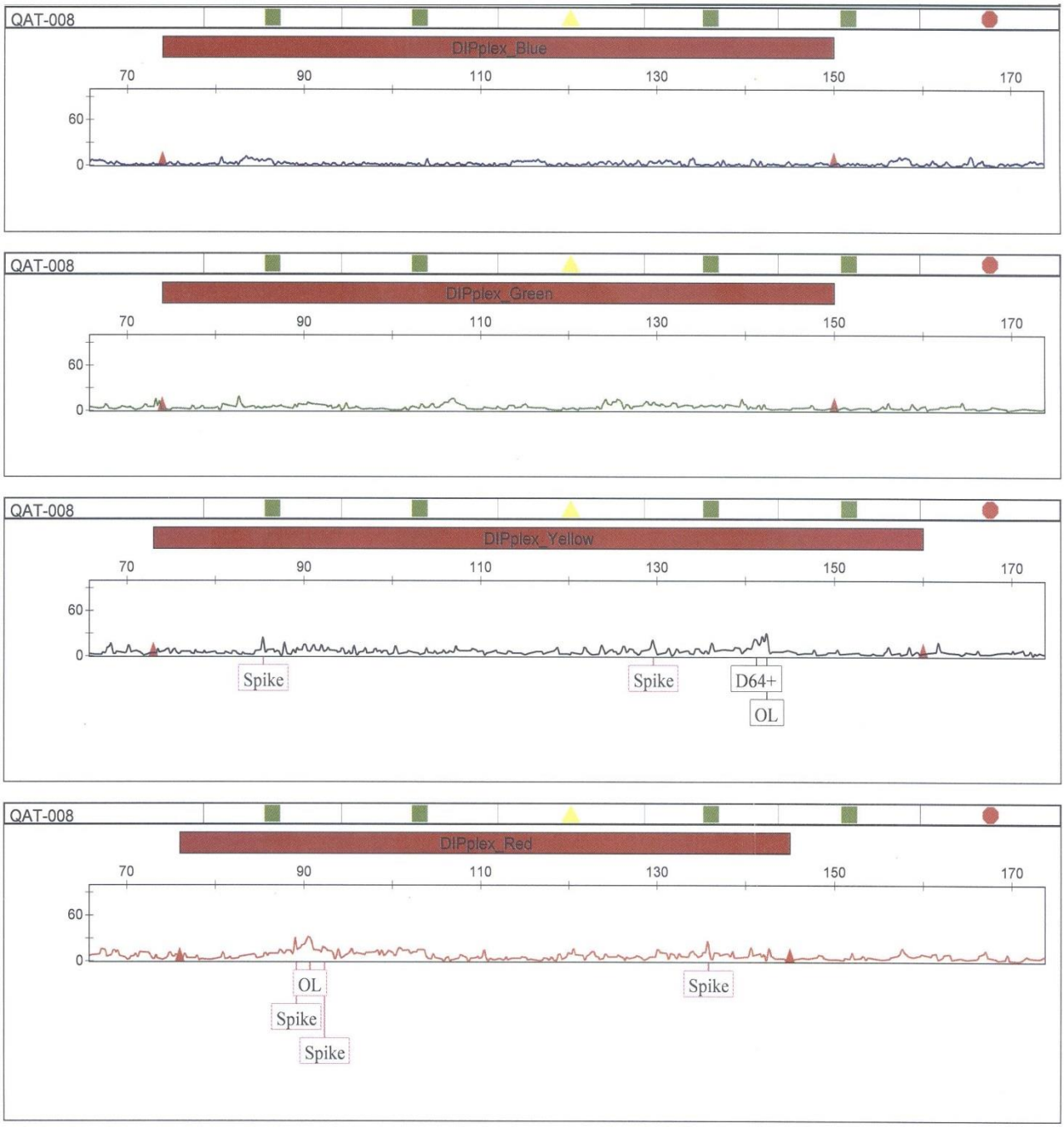

Figure 5.21 Original electropherogram showing an example of the profile generated from a Qatari DNA sample (QAT-008) using Investigator® DIPplex kit (n = 25) that shows no amplification of any of the loci except some artefacts.

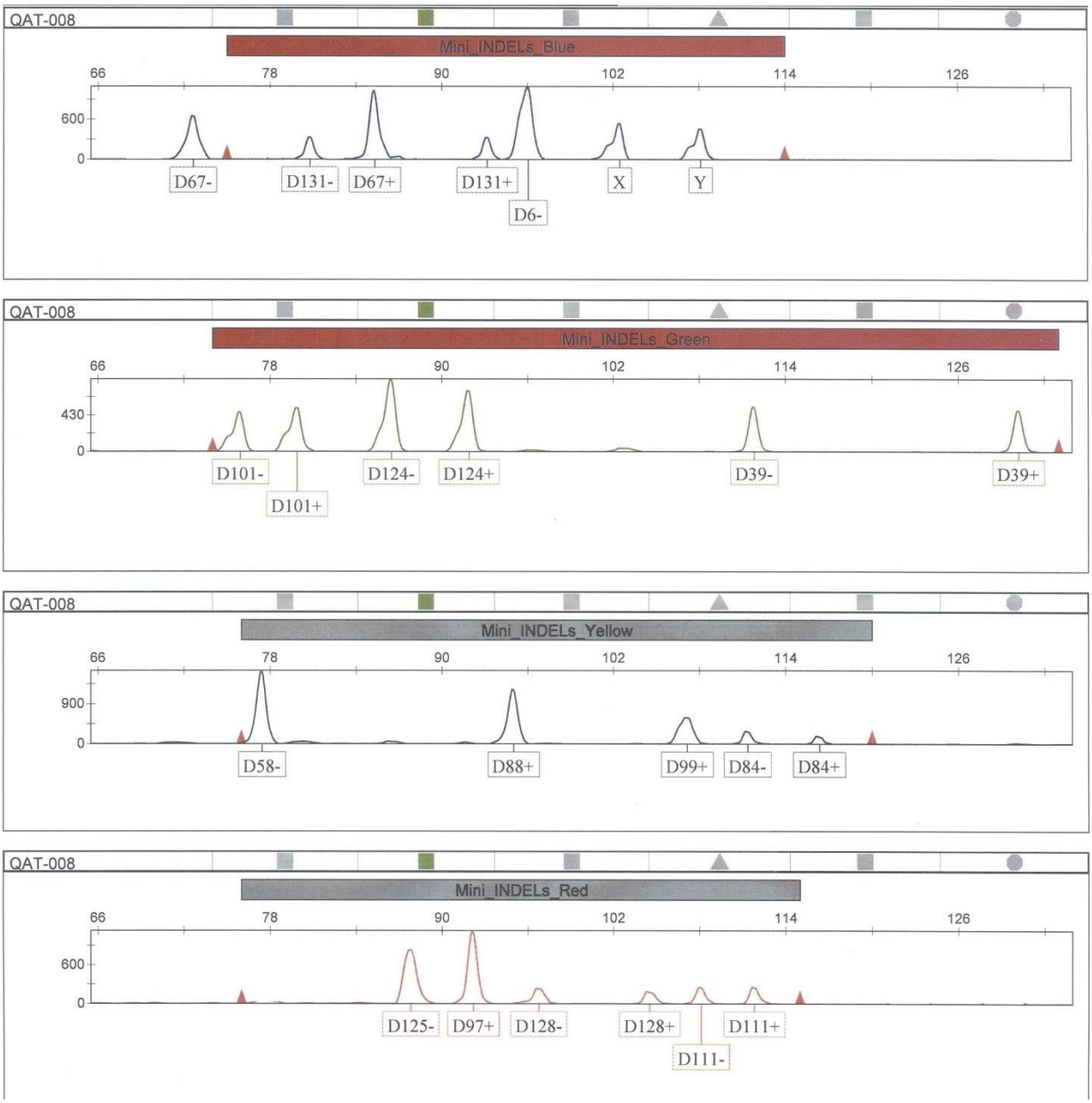

Figure 5.22 Original electropherogram showing an example of the profile from a Qatari DNA sample (QAT-008) generated using mini- INDELs kit (n = 25) that shows full amplification of all of the loci.

5.3.8. Statistical value of 14 mini-INDELs

The statistical values of 14 mini-INDELs were calculated using the data generated during analysis of 30 INDELs from Qiagen DIPplex kit for 500 samples from five different populations i.e. Qatari, Pakistani, Tunisian, Yemeni and Sudanese (as 14 markers are common between two kits).

The forensic suitability of the 14 mini-INDEL loci was evaluated for all the five population groups in Qatar (Table.5.13). The Combined Power of Discrimination (CPD) for the 14 INDEL loci was 0.99999 for all of five populations. The Combined Probability of Match (CPM) was calculated in the range of 1.76×10^{-6} (Tunisian) to 7.77×10^{-7} (Yemeni). The Combined Probability of Exclusion (CPE) was found to be in the range of 0.9367 (Sudanese) to 0.9633 (Pakistani). Table 5.13 represents the values of CPD, CPE and CPM for both kits.

Table 5.13 Table representing the comparison of 15 mini-INDELs with 30 Investigator® DIPPLEX INDELs for five different population groups based in Qatar (n = 500).

	MINI-INDELS (14 Markers)				**DIPPLEX INDELS (30 Markers)**		
	CPD	**CPE**	**CPM**		**CPD**	**CPE**	**CPM**
QATARI	0.999998	0.9420432	$7.651E^{-07}$		0.9999999	0.9973	$1.05E^{-12}$
PAKISTANI	0.999998	0.9633785	$1.153E^{-06}$		0.9999999	0.9974	$8.90E^{-13}$
SUDANESE	0.999938	0.9367869	$3.805E^{-06}$		0.9999999	0.9963	$5.57E^{-12}$
YEMENI	0.999983	0.9378613	$7.770E^{-07}$		0.9999999	0.9963	$7.91E^{-13}$
TUNISIAN	0.999996	0.9550094	$1.764E^{-06}$		0.9999999	0.9964	$9.95E^{-13}$

5.4. DISCUSSION

The main objective of this study was to develop a multiplex assay based on Insertion Deletion polymorphism that can amplify degraded DNA samples and low quantity samples using shorter amplicons. The basic idea for the development of this multiplex was derived from the concept of miniplex (reduced size primer sets), which was introduced to address the issue of DNA degradation (Butler *et al.* 2003).

The sequence data of mini-INDEL markers which were available for different population groups from NCBI website were aligned using BioEdit Sequence Alignment Editor (Hall 2011). The alignment of these sequences of mini-INDELs was helpful to check that flanking region around the point of insertion or/and deletion are same among different population groups. Then, these alignments of sequences helped in designing of primers.

All of 14 mini-INDEL markers were first amplified in singleplex form to determine their primer efficiency and expected amplicon size. The markers were optimised in terms of primer concentration, annealing temperature and input amount of DNA. The best parameters were found to be 0.5 µM of primer concentration, 58 °C annealing temperature and 0.5 ng of input DNA. After the optimisation of mini-INDELs as singleplex, the multiplex of 14 INDELs along with amelogenin was developed successfully in a single tube reaction. All the markers in the multiplex were amplified efficiently with good peak height and expected product sizes. During the analysis of some DNA profiles, it was observed that two alleles, HLD67+ and HLD131- were only one base apart, which resulted in split peak pattern in the presence of both alleles. To address this issue, HLD67 reverse primer was re-designed by adding a tail of 5 bases (TGAAT) to it (Ballard *et al.* 2002). The new reverse primer of HLD67 was re-synthesized and optimised again. Later on, both alleles were separated by 3 bases apart and showed two distinctive peaks. An allelic ladder was successfully developed by combining all the heterozygous alleles for all of 14 INDEL markers (Smith *et al.* 1995; Griffiths *et al.* 1998). The reproducibility and precision of the results from mini-INDELs kit were evaluated by re-analysing the data from different runs to determine the consistency of sizing. The consistency of the sizing was tested by analysing the average size and standard deviation of all alleles for each DNA sample (Hartzell *et al.* 2003).

It is a common observation during case work analysis that sometime DNA concentration is very low even not degraded (Gill *et al*. 2001). DNA profiling of these low quantity samples results in no or little amplification (Schulz *et al*. 2002). In order to recover possible genetic information in this type of scenario, it was presumed that reduced PCR products would improve the chances of amplification success (Butler *et al.* 2003). To test this hypothesis, a sensitivity study was performed in which a set of DNA samples were prepared (1 ng, 0.5 ng, 0.25 ng, 0.125 ng, 0.0625 ng, 0.03125 ng and 0.0156 ng) and then amplified in triplicate with four commercial kits (Applied Biosystems™ Identifiler Plus®, Applied Biosystems™ MiniFiler®, Applied Biosystems™ Y-Filer® and Qiagen™ Investigator® DIPplex) along with mini-INDELs kit.

The results of this sensitivity study indicated that mini-INDELs kit was more sensitive than those of all the other kits. The mini-INDELs kit was sensitive upto 0.03125 ng with complete and balanced profile. The results showed that it amplified more than half of its loci for DNA concentration of 0.0156 ng. The possible reason of successful amplication of low quantity DNA samples is the presence of shorter amplicons of mini-INDEL kit (< 120 bp except D39 + allele which is 130 bp long) while all other kits have larger amplicons. This proved the hypothesis that shorter amplicons are more effective in recovering profiles than those of longer amplicons (Chung *et al.* 2004; Grubwieser *et al.* 2006).

On the other hand, Applied Biosystems™ Identifiler Plus® kit was sensitive upto 0.125 ng which is further confirmed by its sensitivity results generated during its validation. On the other hand, other studies showed that it was capable of producing quality profile upto 0.125 ng using 28 cycles (Collins *et al.* 2003; Wang *et al.* 2011). The sensitivity of Identifiler Plus® kit might be enhanced using increased number of PCR cycles from the recommended ones (28), but the risk of allele drop-out, allele drop-in and contamination would be increased in this case (Schulz *et al.* 2002).

The validation results of Y-Filer® amplification during its development and later on studies showed that it was capable of producing quality profiles for 0.062 ng of DNA (Y-Filer® Manuel; Rodig *et al.* 2008). But the sensitivity study during this research showed different results. The results also showed that Y-Filer® was capable of producing quality profiles for the DNA samples

of 0.25 ng and 0.125 ng. While drop-outs were observed for higher amounts of 1 ng and 0.5 ng and for lesser amounts of 0.625 ng and 0.03125 ng. The possible explanation of this difference in sensitivity results for Y-Filer® might be due to the use of detection instruments. The sensitivity study of Y-Filer® by manufacturer was done using a Genetic Analyser 3100. The sensitivity study in this research was carried out on Genetic Analyser 3500, which is more sensitive in detecting DNA fragments due to its more consistent temperature control, presence of better laser and CCD camera (Hill and Butler 2010). In addition to difference in detection system, Y-Filer® kit in general needs a lesser amount of DNA to amplify as compared to autosomal STRs. Thus, less than or more than optimum amount of input DNA might result in preferential amplification (Walsh *et al.* 1992; Chung *et al.* 2004).

The validation of the manufacturer (Applied Biosystems™) indicated that MiniFiler® kit was sensitive upto 0.062 ng to generate a DNA profile (MiniFiler® manual) and it was also proved in this study. Even it amplified partially 0.0125 ng DNA.

The sensitivity result for Qiagen™ Investigator ®DIPplex kit was not available in its manual, but the recommended input of DNA from manufacturer and in other studies (Akhteruzzaman *et al.* 2013; Nunotani *et al.* 2015), was 0.5 ng. It was also proved in this sensitivity study that 0.5 ng was the optimum amount needed for its amplification to produce a quality profile.

A concordance study was carried out between mini-INDELs kit and Qiagen DIPplex kit using 175 DNA samples from four different populations. The concordance between two kits (for 14 common markers) was observed in 99.7% INDEL alleles calls. A similar concordance (99.7%) was achieved while comparing Applied Biosystems™ Identifiler® and MiniFiler® kits (Hill *et al.* 2007, Mulero *et al.* 2008, Alenizi *et al.* 2009). The possible reason of non-concordance of alleles might be the presence of insertions or deletions in the flanking region outside the primer binding sites (Clayton *et al.* 2004, Hill *et al.* 2007).

Dye blobs were observed during the analysis of some profiles generated from mini-INDELs kit (Figure 5.15). The presence of dye blobs are often observed when the PCR products have shorter amplicons around 120 bp. Dye blobs are usually the result of incomplete attachment of dye molecules during primer synthesis. Dye blobs can easily be distinguished from real peak by

its wider and less intense morphology and can be removed by using filtration columns after PCR reaction (Butler *et al.* 2003).

The use of internal size standard is very critical for accurate genotyping. In this study, LIZ-500 internal size standard (Applied Biosystems™) was used during the genotyping of mini-INDELs. Although, it was recommended to use BTO-550 for the sizing of short amplicons by Investigator® DIPplex kit manual, but the sizing of DNA fragments was more precise and accurate with LIZ-500 than that of BTO-550.

During this study, three Genetic analysers (310, 3130XL and 3500XL) were used for the detection of PCR products. A comparative study revealed after the analysis of the profiles generated from these three instruments that they differ in terms of sensitivity and quality of data. For example, a profile that has been generated from Genetic analyser 310 is of average peak height of 1000 RFU with artefacts while from Genetic analysers 3130XL and 3500, it might have 5000 RFU and 8000 RFU respectively and without any artefacts. The reason of better sensitivity and quality of profiles are attributed to more consistent temperature control, solid state laser (instead of argon ion laser in GA 310) and multi wave length analyser i.e. Charge Coupled Device (CCD) camera of multi capillaries instruments i.e. 3130XL and 3500 (Hill and Butler 2010). Another observation was made regarding the bin settings of INDEL alleles during the comparison of profiles generated from three instruments. There was an average difference of 1-2 base pair of bin setting of allele between the profiles generated from GA 310 and GA 3130XL/3500. For example, if allele D67+ was observed at 84 base pair for the profile generated from GA 310 then it might appear at 86 base pair for the profile generated from GA 3130XL. This problem was resolved by creating different bin settings for each instrument and was used accordingly while analysing the data in GeneMapper ID-X.

A comparative study was also conducted between kits of mini-INDELs and Qiagen™ Investigator® DIPplex using 25 DNA samples. These DNA samples were poorly amplified with Investigator® DIPplex kit while full profiles were generated using mini-INDELs kit by using the same amount of DNA. These results show that mini-INDELs kit is more sensitive than Qiagen™ Investigator® DIPplex kit due to the presence of shorter amplicons (Chung *et. al.* 2004).

The forensic efficiency of mini-INDELs kit was evaluated by calculating Combined Power of Discrimination (CPD), Combined Power of Exclusion (CPE) and Combined Probability of Match (CPM) for five population groups based in Qatar and then compared with those of Investigator® DIPplex kit (Table.5.7). Obviously, due to half number of INDELs markers as compared to Investigator® DIPplex kit, all the values of CPD, CPE and CPM were lower but still they were significant to be used as supplementary kit.

5.5. CONCLUSION

In this study, a new mini-plex of INDELs have been developed which can be used to amplify low quantity and degraded DNA samples. Like other commercially developed multiplexes, this system has the ability to amplify number of markers in a single tube reaction. While comparing to other bi-allelic markers such as SNPs which require additional step for their typing and complex interpretation procedures, mini-INDELs system reduces laboratory procedures and can be easily implemented in the conventional forensic laboratories with same flow of work which is being used for the analysis of conventional STRs.

In conclusion, it can be said confidently that mini-INDELs system can be incorporated in the criminal laboratories as a useful forensic tool and can also be used in conjunction with STRs as a complementary tool especially in those cases where low level of DNA and DNA degradation are suspected.

CHAPTER 6
APPLICATION OF MINI-INDELS TO THE ANALYSIS OF CHALLENGED SAMPLES

6.1. OVERVIEW

Most of the forensic DNA samples collected from crime scene are typically less than ideal. They may have been exposed to a harsh environment for longer periods, such as in homicide cases. As a result, the DNA samples might be degraded or/and inhibited severely. Due to the deterioration of DNA, forensic scientists can find it difficult to successfully process these challenged samples. The forensic DNA profiling of these degraded and inhibited DNA samples can result in loss of signal, allele drop-out and imbalance alleles (Chung *et al.* 2004).

The inhibited DNA samples often produce partial profile with drop outs of larger amplicons that look similar to a degraded DNA sample. Thus, failure of larger loci to amplify can be either due to DNA degradation (not enough DNA templates) or due to PCR inhibition (preventing the polymerase activity). The use of smaller amplicons (such as mini-INDELs) might aid in recovery of genetic information in both of these cases (Opel *et al.* 2008). In this Chapter, an effort was made to investigate the application of mini-INDELs (reduced size amplicons) on a set of artificially prepared degraded and inhibited samples.

6.1.3 Objectives

This Chapter has the following objectives:

(a) To determine the rate of DNA degradation in blood samples in relation to ADD °C following environmental insult.

(b) To evaluate the efficiency of mini-INDELs amplification kit by applying it to the set of challenged samples (degraded and inhibited) prepared under controlled conditions.

(c) To compare the efficiency of mini-INDELs kit with that of commonly used commercial STR kits (Identifiler Plus® and MiniFiler® from Applied Biosystems™) by using same set of artificially degraded and inhibited DNA samples.

6.2. Methods and Materials

6.2.1 Preparation of degraded DNA samples by using environmental insults

In this study, 5 µl and 10 µl of fresh blood were deposited on cloth and metal surfaces respectively. Samples were placed outdoor for the duration (3 months) of the experiment during the months of June, July and August and the temperature was measured in terms of accumulated degree-days (ADD °C). Samples were collected every 3-6 days and extracted using the phenol chloroform isoamyl protocol described in Chapter 2.

Extracted DNA samples were subjected to quantification by using Human Quantifiler® Trio according to the manufacturer's recommended protocols (Applied Biosystems™).

6.2.2 Preparation of degraded DNA samples by chemical means

Chemical degradation of DNA was achieved by adding different amounts of NaClO (bleach) and H_2O_2 (Hydrogen Peroxide) to 1 ng of control DNA 9947A (Table 6.1).

Table 6.1 Table showing the two selected reagents used to induce chemical degradation along with their concentrations, which were added to control DNA 9947A used for experiments.

Degradation agents	Concentration (Percentage)
NaClO (bleach)	0.15,0.3,0.4,0.5,0.6, 0.8
H_2O_2 (Hydrogen peroxide)	0.1,0.2,0.3,0.4, 0.6

6.2.3 Preparation of samples by inducing PCR inhibitions

PCR inhibited samples were prepared by using ethanol and phenol as inhibiting agents. Different concentrations of both reagents were added to 1 ng of control DNA 9947A (Table 6.2).

Table 6.2 Table showing the two selected inhibiting agents along with their concentrations that were added to induce inhibition to control DNA 9947A used for experiments.

Inhibiting agents	Concentration (Percentage)
Ethanol	5.0,7.5,8.5,10
Phenol	0.07,0.09,0.11,0.16,0.25

The downstream procedures which include extraction of DNA samples, quantification, amplification and fragment analysis have been described in Methods and Materials (Chapter 2).

6.3 RESULTS

6.3.1 Effect of DNA degradation using environmental insults

Based on the quantification results (Figure 6.1 and Table 6.3), by using Applied Biosystems™ Human Quantifiler® Trio, the DNA samples were selected (collected after 43, 58, 68, 76, 80 and 84 days) and subjected to amplification. A volume of 10 µl of DNA template from each selected sample was used in total PCR reaction volume of 25 µl. The amplification on the set of degraded samples was performed using Applied Biosystems™ Identifiler Plus®, Applied Biosystems™ MiniFiler® and Mini-INDELs kits. The summary of the amplification results from all of the three kits is shown in Table 6.4. Figures 6.2-6.4 show the electropherograms of the amplification results of three kits and in each electropherogram, same RFU values (at Y-axis) are selected to clearly understand the amplification efficiency (like in Figure 6.2 for Identifiler Plus ® result, 410 RFU is selected for all electropherograms).

The results from the Identifiler Plus® amplification indicated that DNA samples subjected to environmental insults for 43 and 58 days were able to generate complete profiles with good quality and peak height. Moreover, weak amplification with drop-outs at larger loci was observed for the sample for 68 days. While samples from 76, 80 and 84 days were poorly amplified with almost all of alleles dropping out (Figure 6.2 and Table 6.4).

The MiniFiler® kit was able to amplify effectively all the selected DNA samples without any drop-outs and good peak heights except for 84 days which has drop out of two alleles from the total of 9 alleles (Figure 6.3 and Table 6.4).

Mini-INDELs kit was also capable of amplifying all of samples efficiently except DNA sample exposed for 84 days, which has 3 drop outs of total 20 alleles (Figure 6.4 and Table 6.4).

Table 6.3 Table Indicating the details of the experiment designed to demonstrate the effect of environmental insults on the DNA samples including the dates of collection, number of days subjected to environmental insults, accumulated degree days (ADD° C) and quantification results from DNA samples deposited on metal (10 µl of blood) and cloth (5 µl of blood). Shaded portions indicate the selection of those DNA samples which were used for downstream applications i.e. amplification and fragment analysis.

Collection Dates	Number of DAYS	ADD° C	Quantification (ng) Metal	Quantification (ng) Cloth
17/06/15	0	42	12. 10	7.49
20/06/15	3	171	9.16	5.12
23/06/15	6	296	8.41	2.77
26/06/15	9	421	7.65	1.74
29/06/15	12	544	7.03	1.39
02/07/15	15	661	6.62	1.30
05/07/15	18	780	5.53	1.03
08/07/15	21	915	4.15	0.82
11/07/15	24	1045	3.67	0.67
14/07/15	27	1175	3.07	0.62
17/07/15	30	1303	2.69	0.39
23/07/15	36	1565	2.05	0.18
30/07/15	43	1862	1.95	0.10
03/08/15	47	2039	1.63	0.09
06/08/15	50	2166	0.97	0.07
11/08/15	55	2374	0.64	0.04
14/08/15	58	2497	0.22	0.03
18/08/15	62	2653	0.03	0.02
24/08/15	68	2909	0.01	0.02
01/09/15	76	3256	0.00	0.017
05/09/15	80	3439	0.00	0.010
09/09/15	84	3598	0.00	0.008
15/09/15	90	3860	0.00	0.00

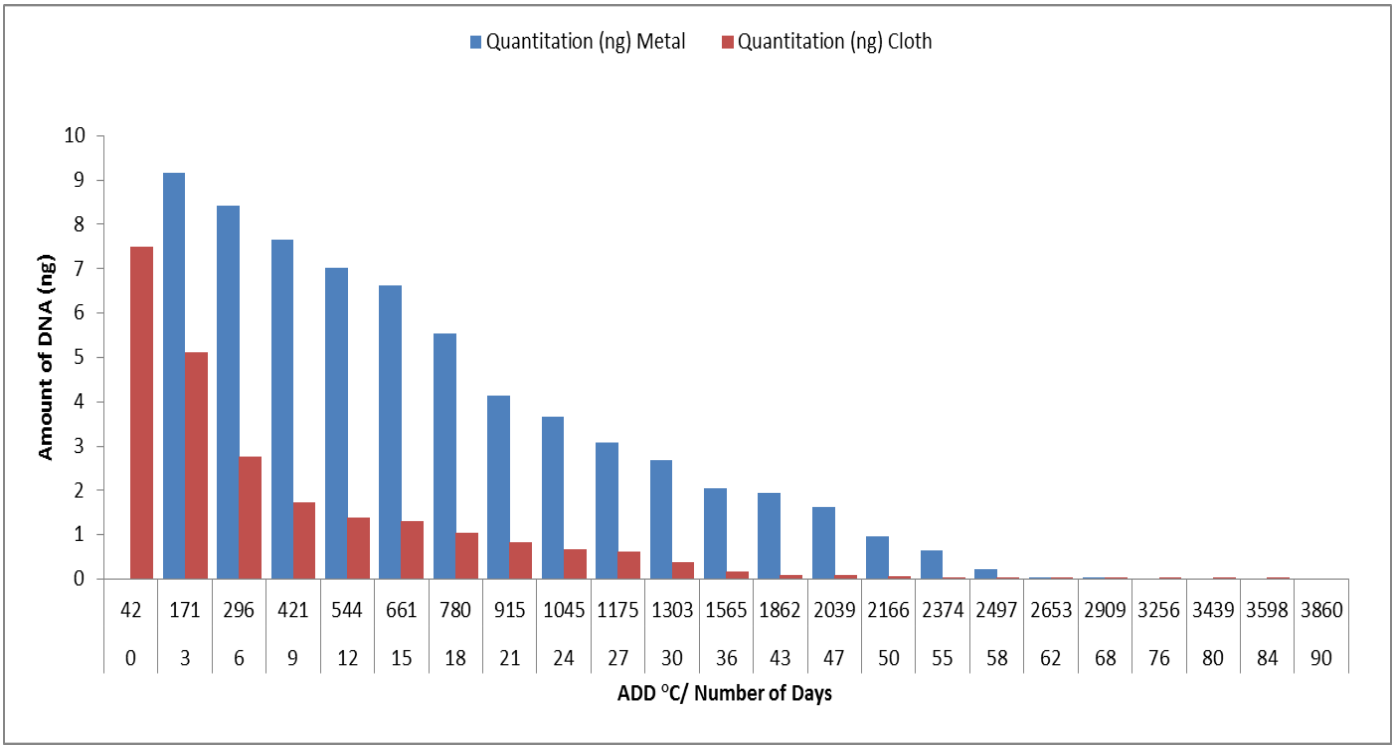

Figure 6.1 Bar diagram Indicating the quantification of DNA samples subjected to environmental insult using the substrates of metal and cloth in relation to accumulated degree days (ADD °C) and number of days.

Table 6.4 Table showing the effects of DNA degradation in relation to ADD ° C on the amplification results of Identifiler Plus® (ID+), MiniFiler® (MF) and mini-INDELs (MINI) kits. The number of drop-outs from total number of alleles and average peak heights are also shown.

# of Days	DNA concentration (ng)	# of alleles drop-outs			Average Peak Heights (RFU)		
		ID+	MF	MINI	ID+	MF	MINI
43	0.10	0/29	0/9	0/20	350	1500	2500
58	0.03	01/29	0/9	0/20	150	1000	1500
68	0.02	12/29	0/9	0/20	75	500	1000
76	0.017	25/29	0/9	0/20	50	250	500
80	0.010	26/29	0/9	0/20	<50	125	300
84	0.008	29/29	2/9	3/20	<50	75	100

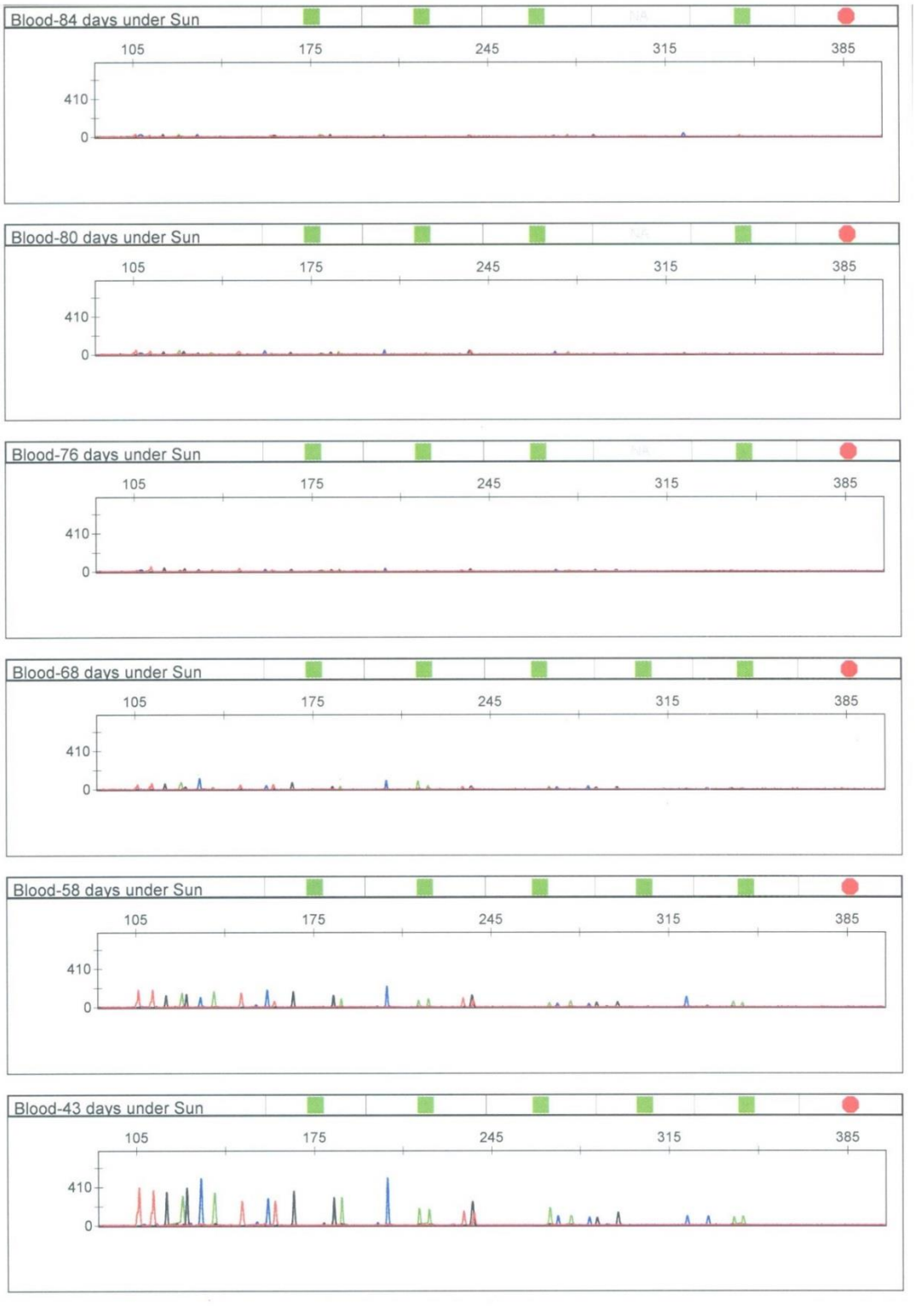

Figure 6.2 Original electropherograms generated from selected DNA samples using 10 µl (subjected to environmental insults) using Applied Biosystems™ Identifiler Plus® kit, (n = 3).

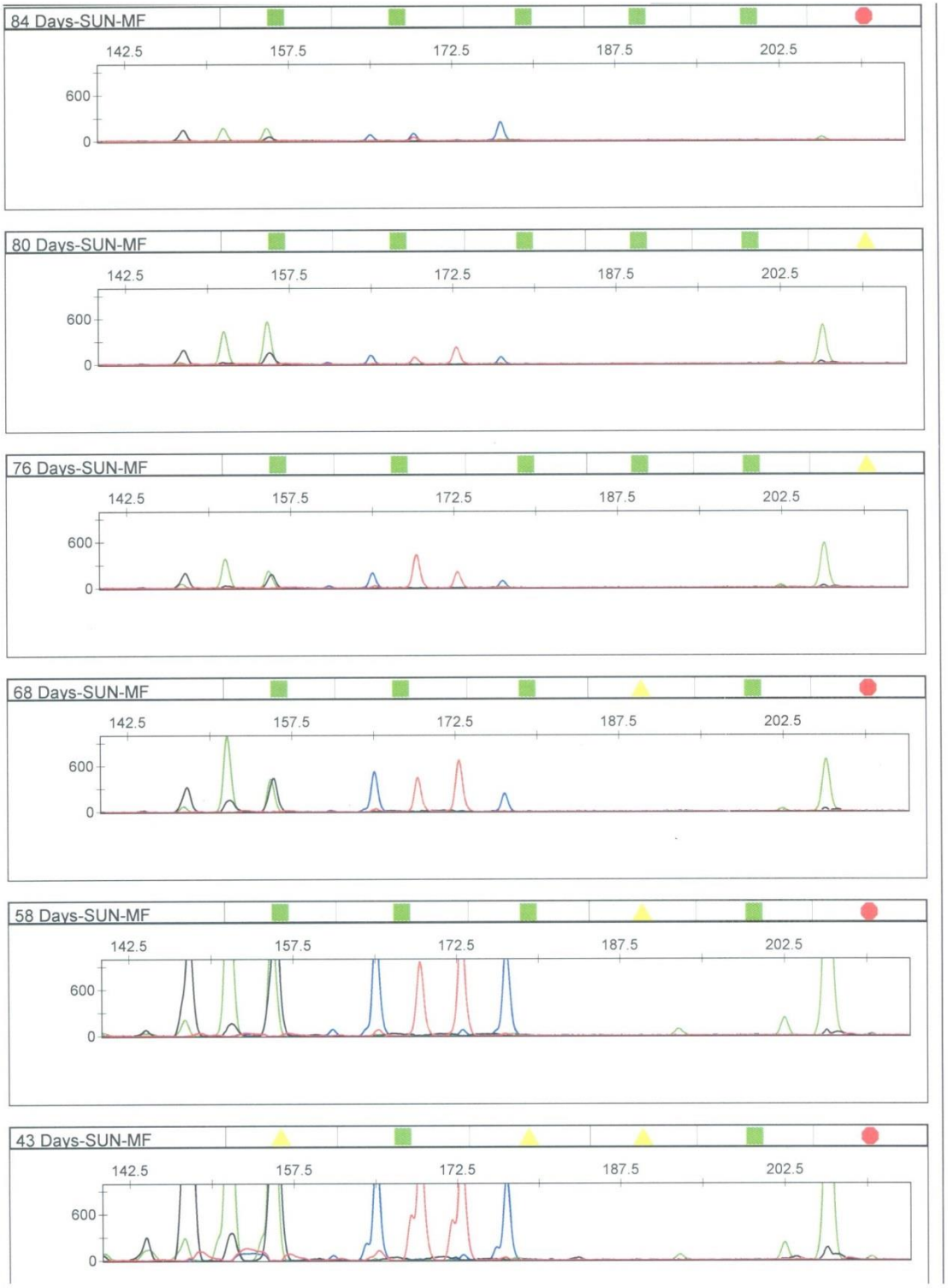

Figure 6.3 Original electropherograms generated from selected DNA samples using 10 µl volume (subjected to environmental insults) using Applied Biosystems™ MiniFiler® kit, (n = 3).

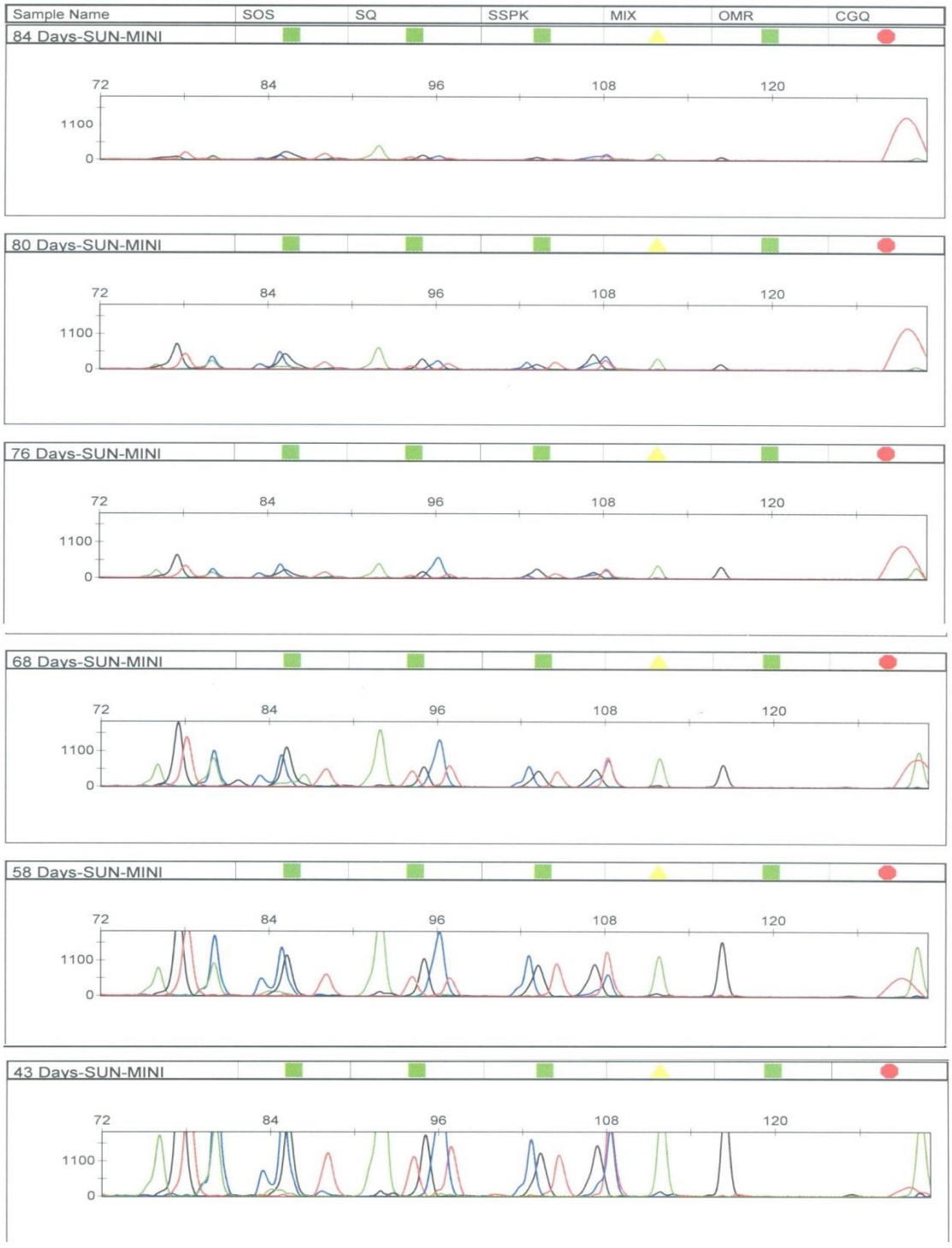

Figure 6.4 Original electropherograms generated from selected DNA samples using 10 µl volume (subjected to environmental insults) using mini-INDELs kit, (n = 3).

6.3.2 Effect of chemical degradation by using Bleach

Bleach was introduced to a set of control DNA (9947A) samples with varying percentages of 0.15, 0.3, 0.4, 0.5, 0.6 and 0.8 respectively for 15 minutes. The amplification was performed using PCR kits of Applied Biosystems™ Identifiler Plus®, Applied Biosystems™ Mini filer® and mini-INDELs.

The results from the Identifiler Plus® amplification shows that DNA samples subjected to bleach concentration from 0.15% to 0.5% were able to generate complete profiles with reasonable peak heights and without any drop-outs. While DNA samples containing 0.6% and 0.8% bleach, were unable to amplify at all (Figure 6.5 and Table 6.5).

MiniFiler® kit was able to amplify effectively complete set of DNA samples with added different concentration of bleach without any drop-outs (Figure 6.6 and Table 6.5).

Mini-INDELs kit also amplifies all of the bleach added DNA samples except with that of 0.8% which has large number of allele drop-outs (Figure 6.7 and Table 6.5).

Table 6.5 Table showing the effects of chemical degradation (by using bleach) on the amplification results of Identifiler Plus (ID+), MiniFiler (MF) and mini-INDELs (MINI) kits. The number of drop-outs from total number of alleles and average peak heights are also shown.

Bleach concentration added to PCR reaction (Percentage)	# of alleles drop-outs			Average Peak Heights (RFU)		
	ID+	MF	MINI	ID+	MF	MINI
0.15	0/29	-	-	900	-	-
0. 3	0/29	-	-	700	-	-
0.4	0/29	-	-	500	-	-
0.5	0/29	0/9	0/20	250	3000	1000
0.6	29/29	0/9	0/20	<50	1500	800
0. 8	29/29	0/9	16/20	<50	700	300

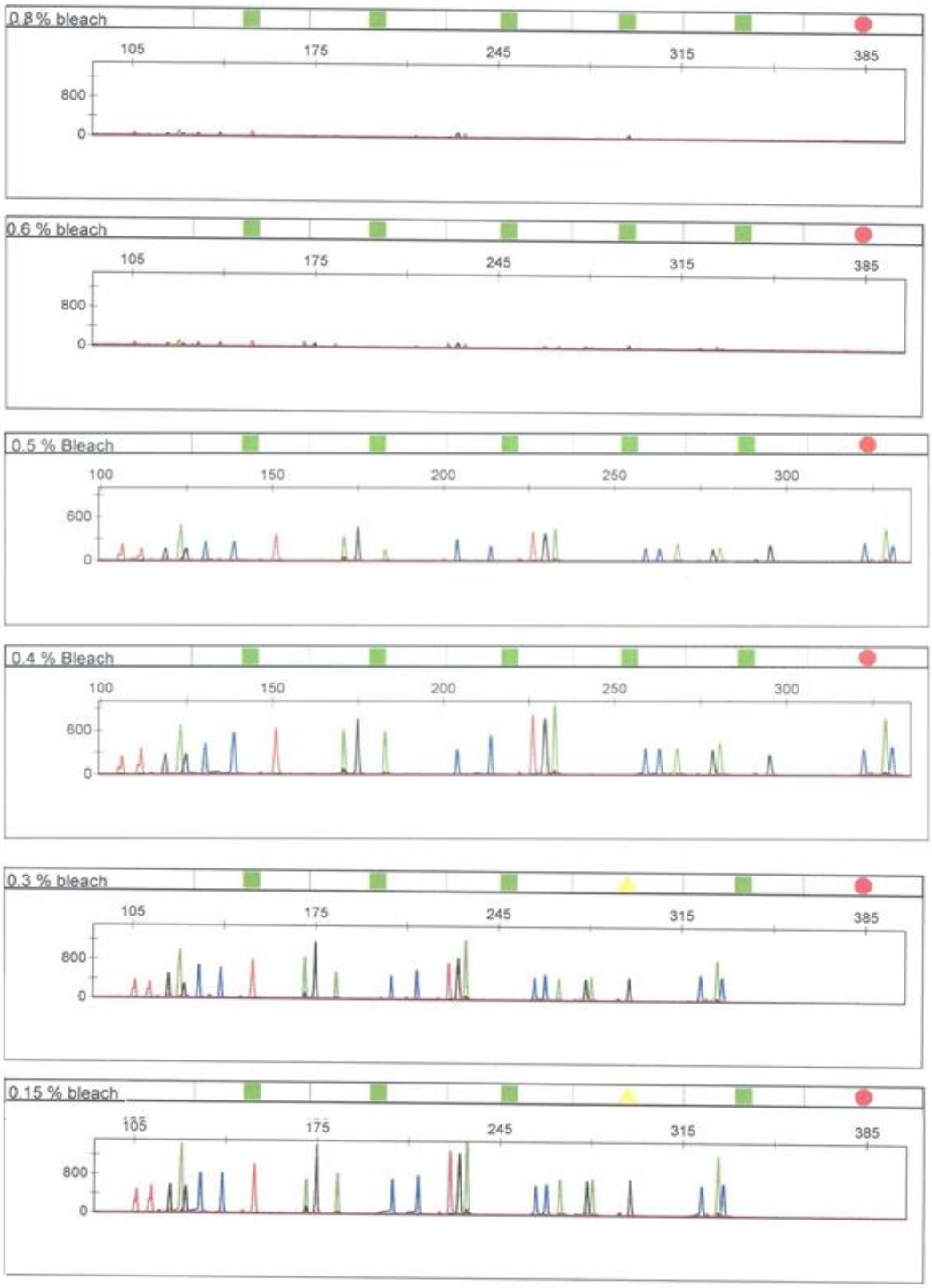

Figure 6.5 Original electropherograms generated from DNA samples (after adding different concentrations of bleach) using Applied Biosystems™ Identifiler Plus® kit, (n = 3).

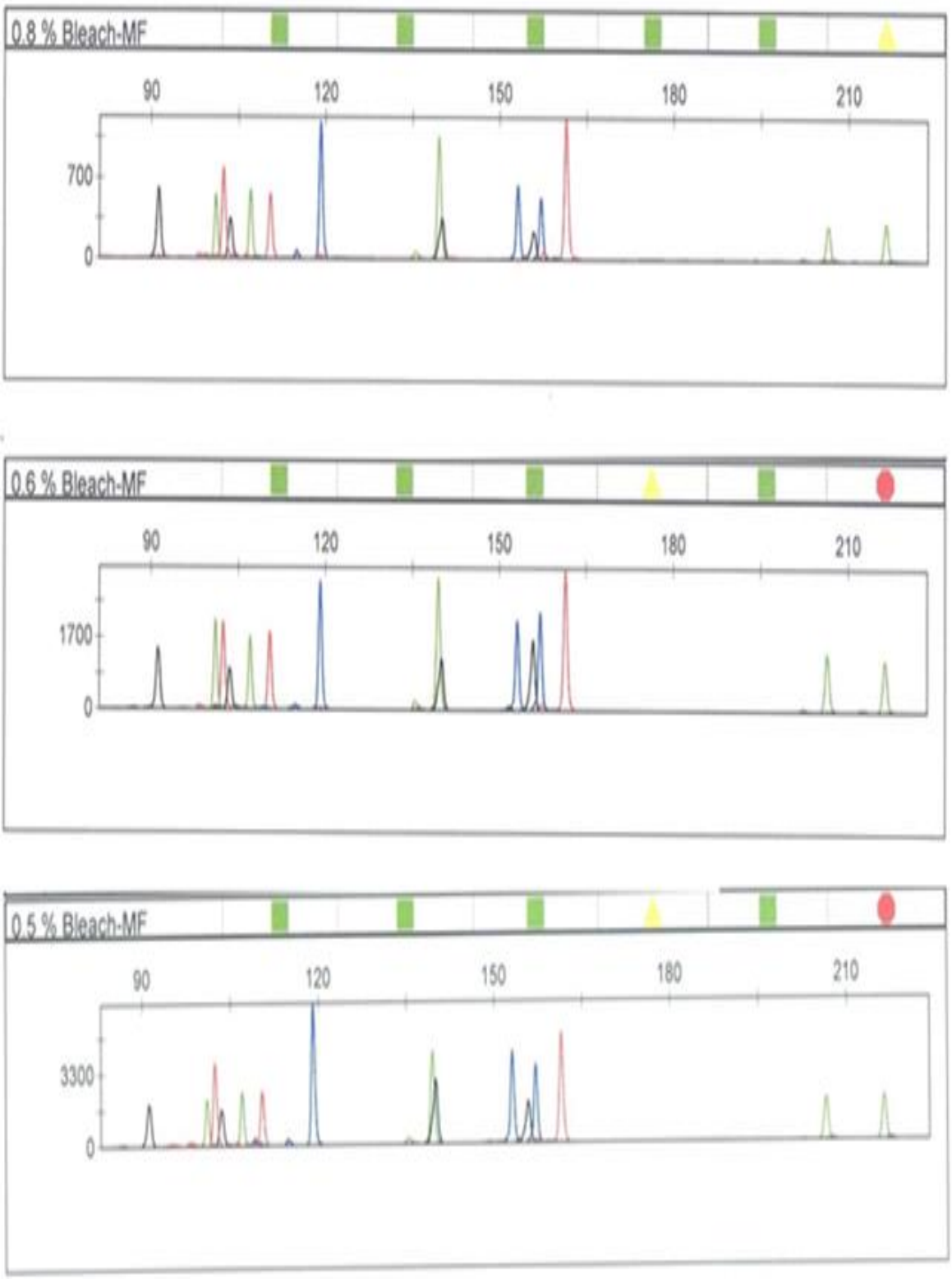

Figure 6.6 Original electropherograms generated from DNA samples (after adding different concentrations of bleach) using Applied Biosystems™ MiniFiler® kit, (n = 3).

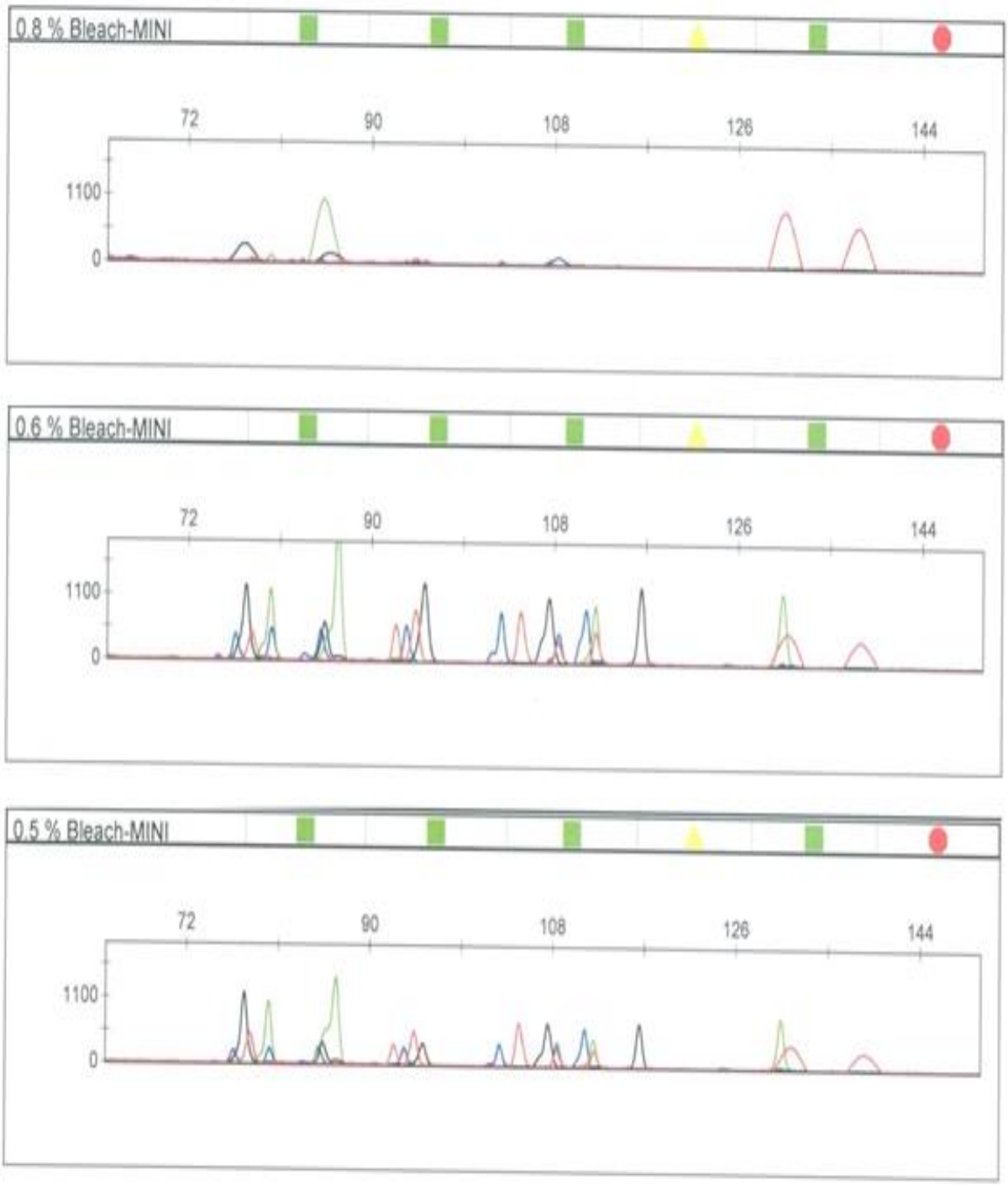

Figure 6.7 Original electropherograms generated from DNA samples (after adding different concentrations of bleach) using mini-INDELs kit, (n = 3).

6.3.3 Effect of chemical degradation by using Hydrogen Peroxide

Another set of DNA samples were prepared to study chemical degradation which was achieved by introducing different concentrations (0.1%, 0.2%, 0.3%, 0.4% and 0.6%) of hydrogen peroxide to a set of control DNA (9947A) for 15 minutes. The PCR kits of Applied Biosystems™ Identifiler Plus®, Applied Biosystems™ MiniFiler® and Mini-INDELs were employed to amplify this set of challenged DNA samples.

The results from the Identifiler Plus® amplification shows that DNA samples subjected to hydrogen peroxide concentration of 0.1% were able to amplify partially (about 50% of loci amplified), while samples with concentration more than 0.1% were unable to amplify at all (Figure 6.8 and Table 6.6).

MiniFiler® kit was able to amplify all of its loci without any drop-outs while using hydrogen peroxide concentration of 0.1%. While for DNA sample with concentration of 0.2%, weak amplification with a single drop-out and with sloping pattern of alleles from shorter to longer amplicons was observed. While the samples with concentration more than 0.2% could not amplify at all (Figure 6.9 and Table 6.6).

Mini-INDELs kit was able to amplify only samples with 0.1% hydrogen peroxide. While other samples with higher concentrations were unable to amplify at all (Figure 6.10 and Table 6.6).

Table 6.6 Table showing the effects of chemical degradation (using hydrogen peroxide) on the amplification results of Identifiler Plus® (ID+), MiniFiler® (MF) and Mini-INDELs (MINI) kits. The numbers of alleles drop-out and average peak heights are also shown.

H_2O_2 concentration added to PCR reaction (Percentage)	# of alleles drop-outs			Average Peak Heights (RFU)		
	ID+	MF	MINI	ID+	MF	MINI
0.1	18/29	0/9	0/20	100	2000	250
0. 2	-	1/9	20/20	-	800	<50
0.3	28/29	6/9	20/20	50	50	<50
0.4	-	9/9	20/20	-	<50	<50
0.6	29/29	9/9	20/20	<50	<50	<50

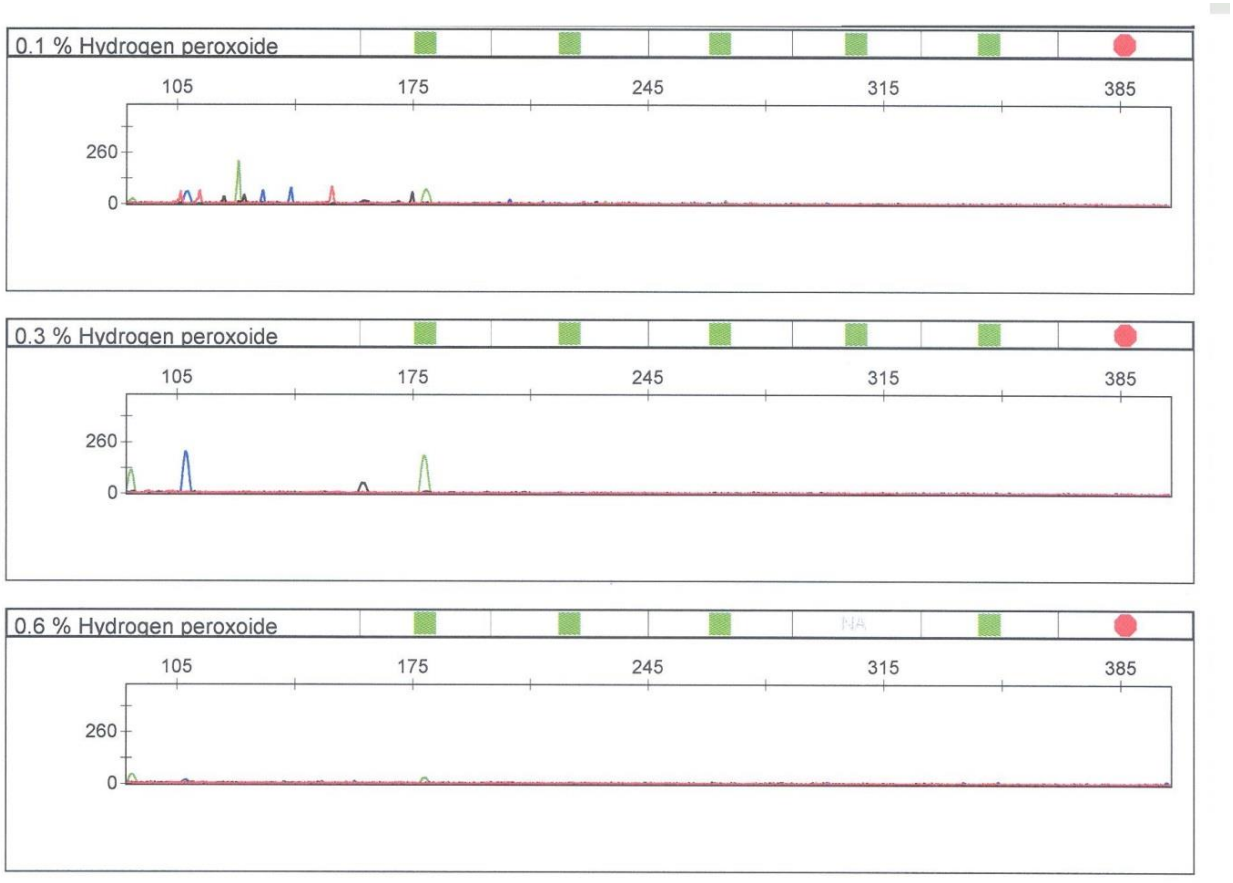

Figure 6.8 Original electropherograms generated from DNA samples (after adding different concentrations hydrogen peroxide) using Applied Biosystems™ Identifiler Plus® kit, (n = 3).

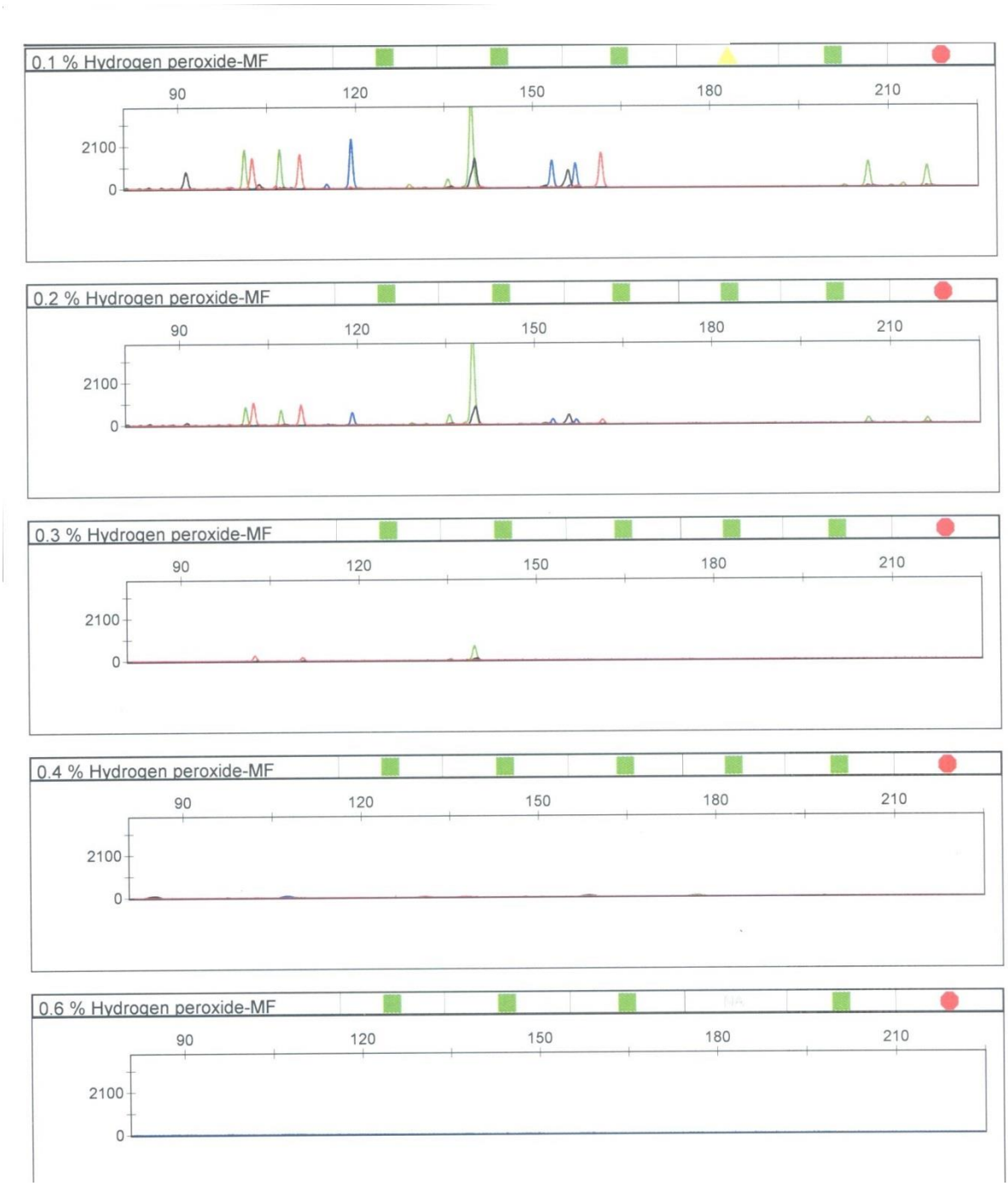

Figure 6.9 Original electropherograms generated from DNA samples (after adding different concentrations Hydrogen peroxide) using Applied Biosystems™ MiniFiler® kit, (n = 3).

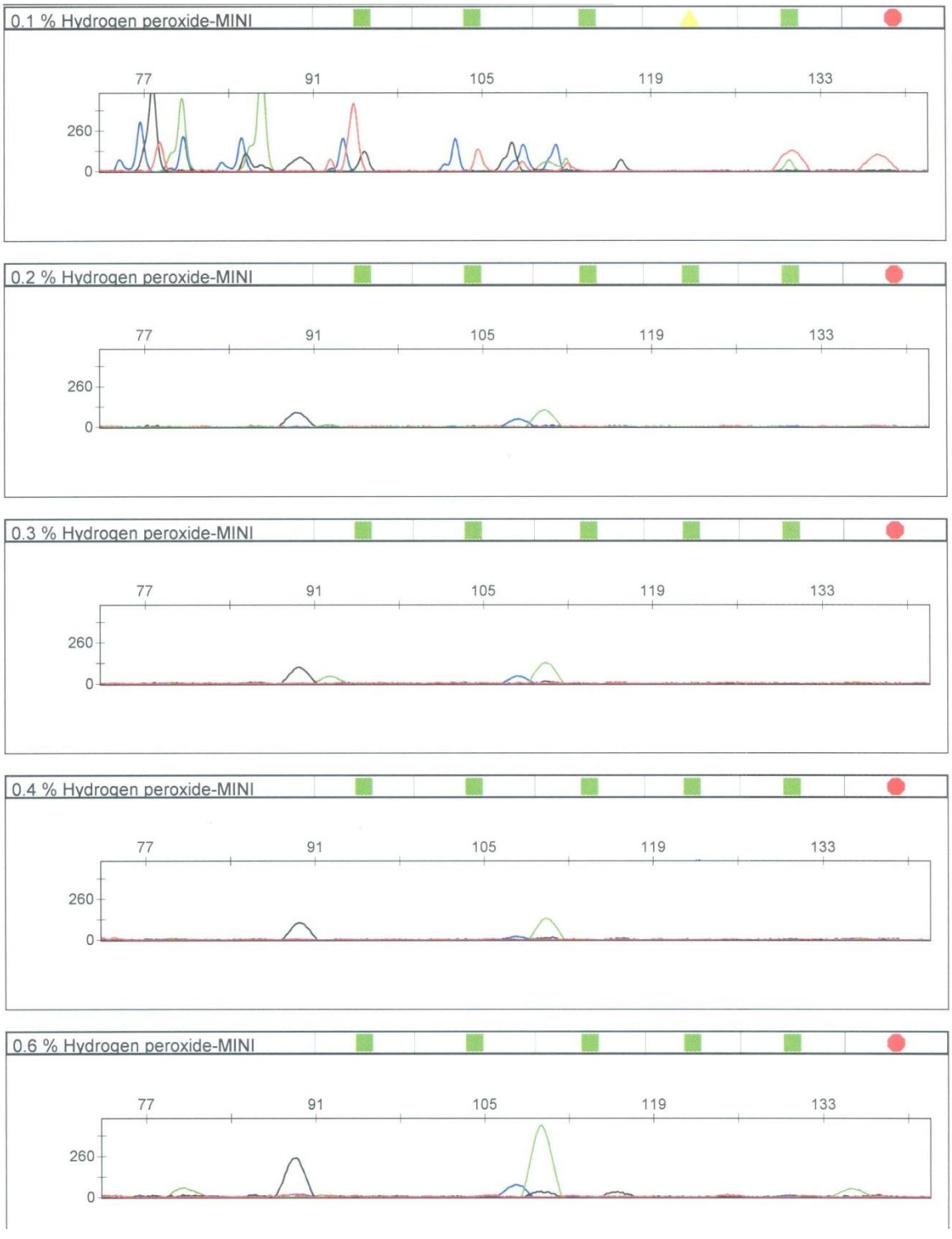

Figure 6.10 Original electropherograms generated from DNA samples (after adding different concentrations Hydrogen peroxide) using mini-INDELs kit, (n = 3).

6.3.4 Effect of Inhibition using ethanol

Ethanol is a commonly encountered PCR inhibitor which in case of its presence in elution results in no or little quantitation of DNA. During this experiment, ethanol was introduced in different concentrations (5.0%, 7.5%, 8.5% and 10%) to a set of control DNA (9947A) samples. Later on, these samples were amplified using PCR assays of Identifiler Plus®, Mini filer® and mini INDELs.

The results from all of three tested kits (Identifiler Plus®, Mini filer® and mini-INDELs) were quite similar while amplifying ethanol inhibited DNA samples. All of three PCR assays were able to amplify only sample with 5% ethanol and for other higher concentrations, they could not amplify even a single locus (Table 6.7 and Figures 6.11-6.13).

Table 6.7 Table showing the effects of inhibition (using ethanol) on the amplification results of Identifiler Plus® (ID+), MiniFiler® (MF) and mini-INDELs (MINI) kits. The number of drop outs from total number of alleles and average peak heights are also shown.

Ethanol concentration added to PCR reaction (Percentage)	# of alleles drop-outs			Average Peak Heights (RFU)		
	ID+	MF	MINI	ID+	MF	MINI
5.0	0/29	0/9	0/20	1000	1000	1000
7.5	29/29	9/9	20/20	<50	<50	<50
8.5	29/29	9/9	20/20	<50	<50	<50
10	29/29	9/9	20/20	<50	<50	<50

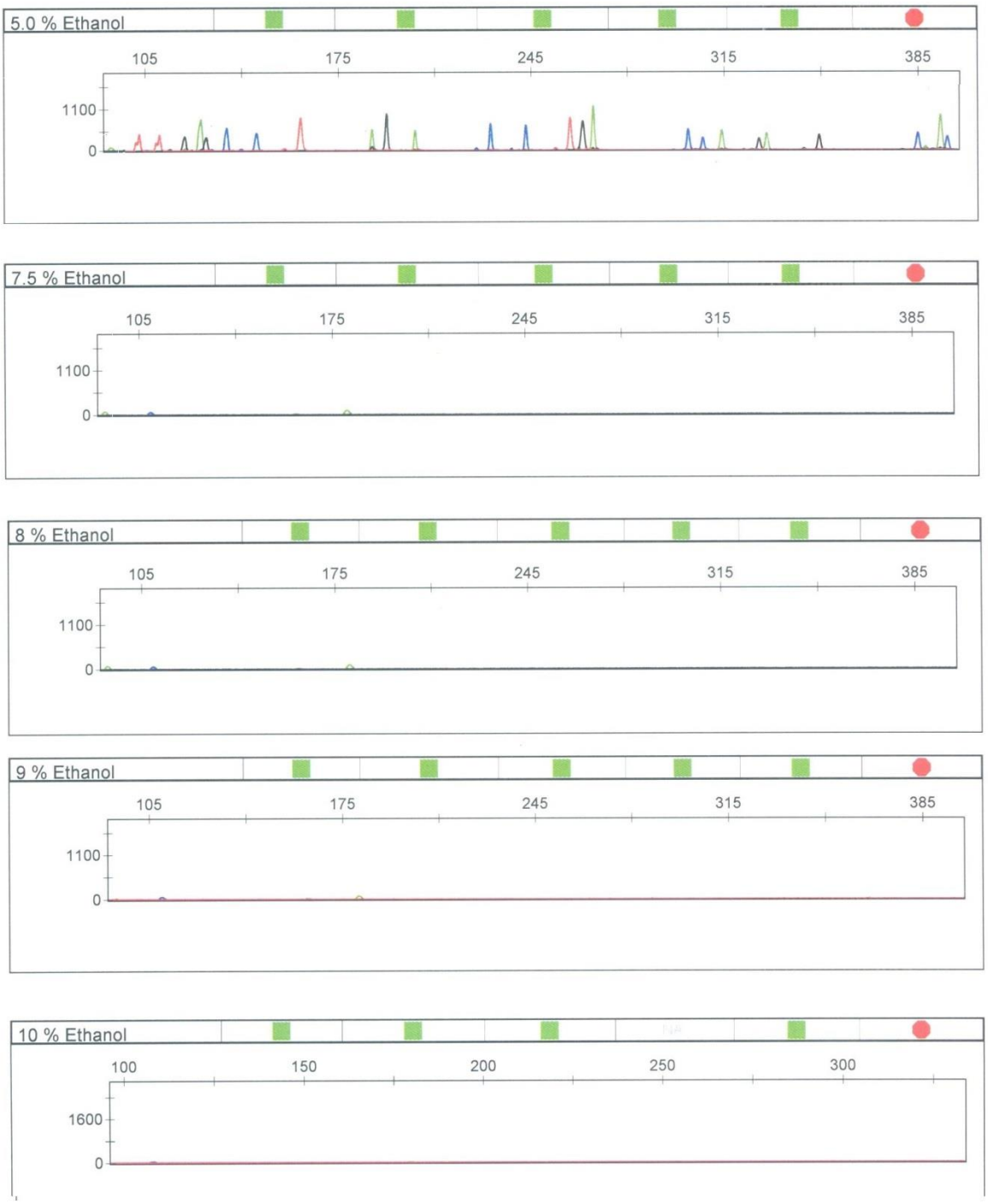

Figure 6.11 Original electropherograms generated from DNA samples (after adding different concentrations of ethanol) using Applied Biosystems™ Identifiler Plus® kit, (n = 3).

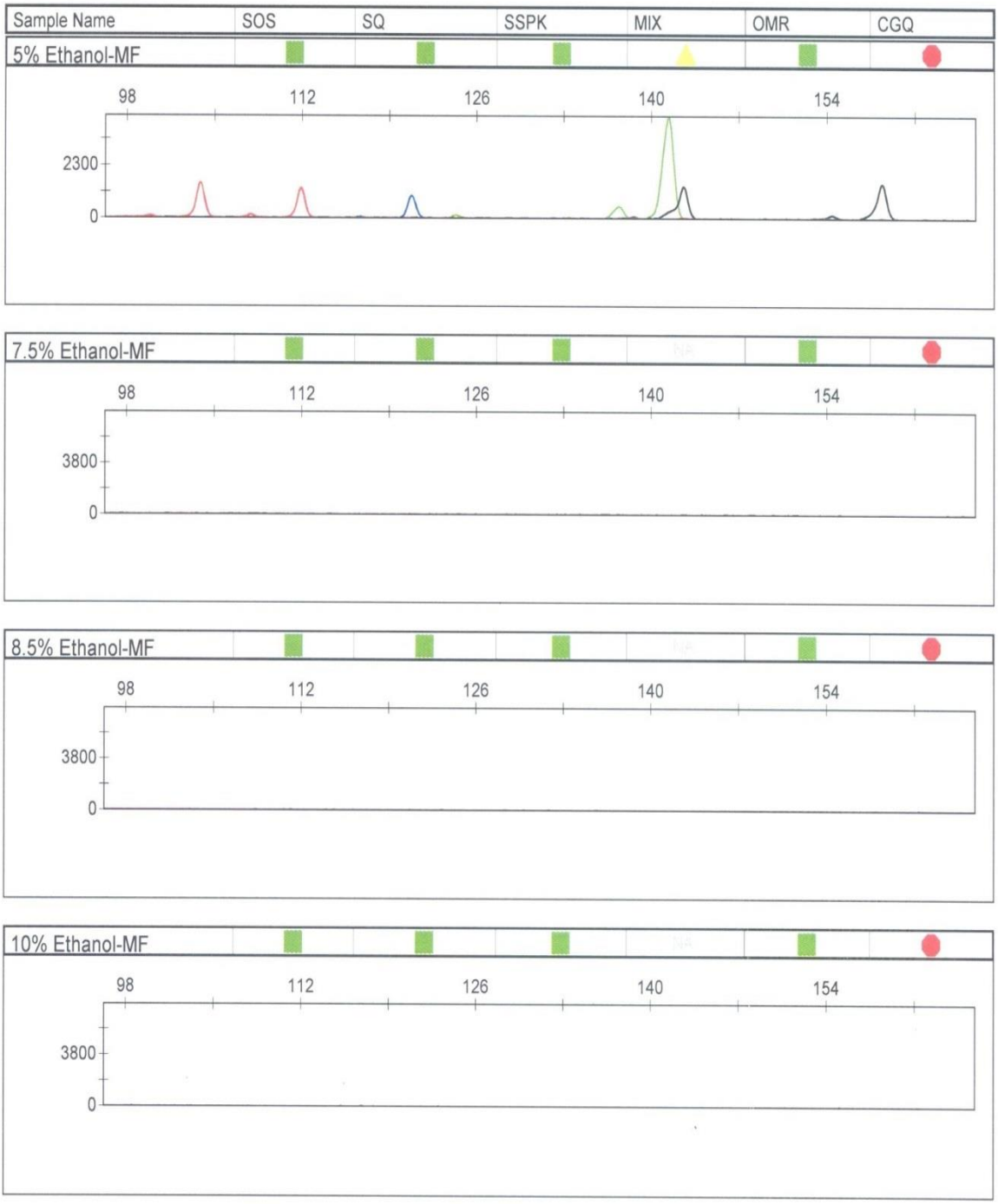

Figure 6.12 Original electropherograms generated from DNA samples (after adding different concentrations of ethanol) using Applied Biosystems™ MiniFiler®, (n = 3).

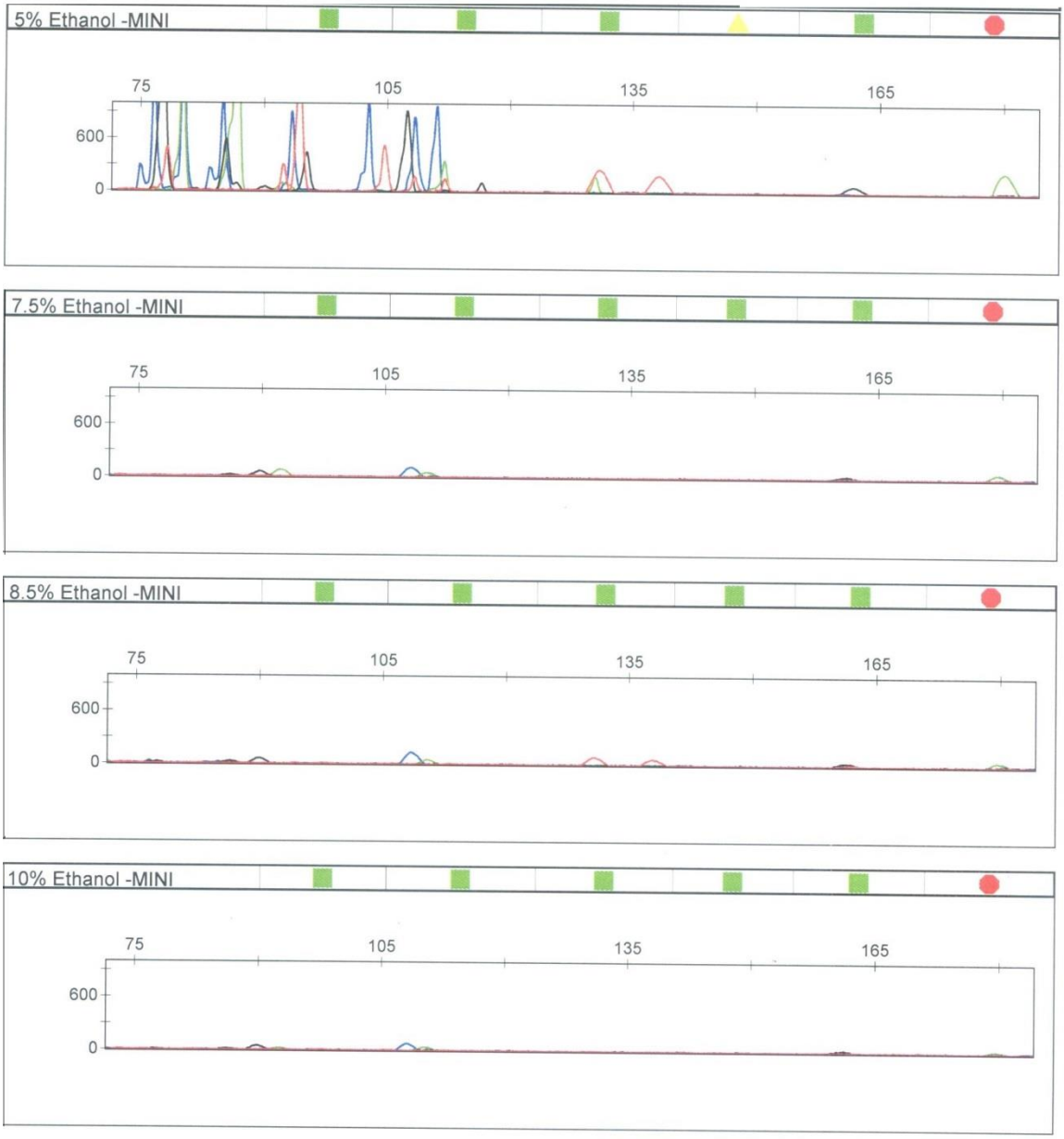

Figure 6.13 Original electropherograms generated from DNA samples (after adding different concentrations of ethanol) using mini-INDELs, (n = 3).

6.3.4 Effect of Inhibition using Phenol

Phenol is an important ingredient of organic based DNA extraction. Its presence in the eluted DNA may result in no amplification. During this experiment, ethanol was introduced in different concentrations (0.07%, 0.09%, 0.11%, 0.16% and 0.25%) to a set of control DNA (9947A) samples. Later on, these samples were amplified using PCR kits of Identifiler Plus®, MiniFiler® and mini-INDELs.

The results from the Identifiler Plus® amplification showed that DNA samples subjected to 0.07%, 0.09% and 0.11% phenol were able to generate quality profiles. While DNA samples with more than 0.11% concentration were unable to amplify at all (Figure 6.14 and Table 6.8).

MiniFiler® kit was unable to amplify any of DNA samples contaminated with phenol (Figure 6.15 and Table 6.8).

While mini-INDELs kit showed a weak amplification of all the samples containing phenol with low peak heights and number of drop-outs (Figure 6.16 and Table 6.8).

Table 6.8 Table showing the effects of inhibition (using phenol) on the amplification results of Identifiler Plus® (ID+), MiniFiler® (MF) and mini-INDELs (MINI) kits. The number of drop-outs from total number of alleles and average peak heights are also shown.

Phenol concentration added to PCR reaction (Percentage)	# of alleles drop-outs			Average Peak Heights (RFU)		
	ID+	MF	MINI	ID+	MF	MINI
0.07	0/29	9/9	14/20	800	<50	80
0.09	0/29	9/9	14/20	650	<50	65
0.11	0/29	9/9	14/20	500	<50	60
0.16	29/29	9/9	14/20	<50	<50	55
0.25	29/29	9/9	14/20	<50	<50	50

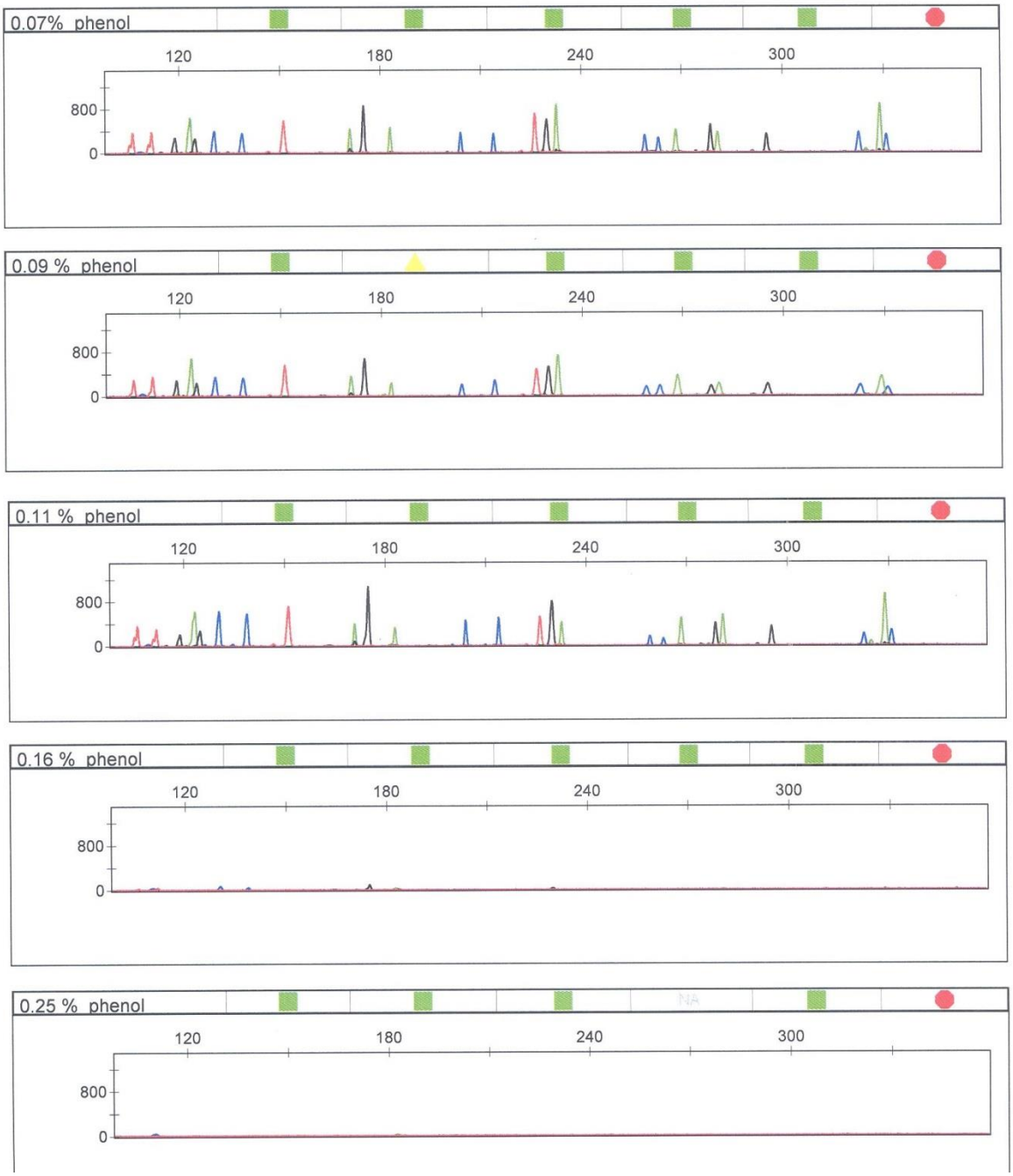

Figure 6.14 Original electropherograms generated from DNA samples (after adding different concentrations of phenol) using Applied Biosystems™ Identifiler Plus® kit, (n = 3).

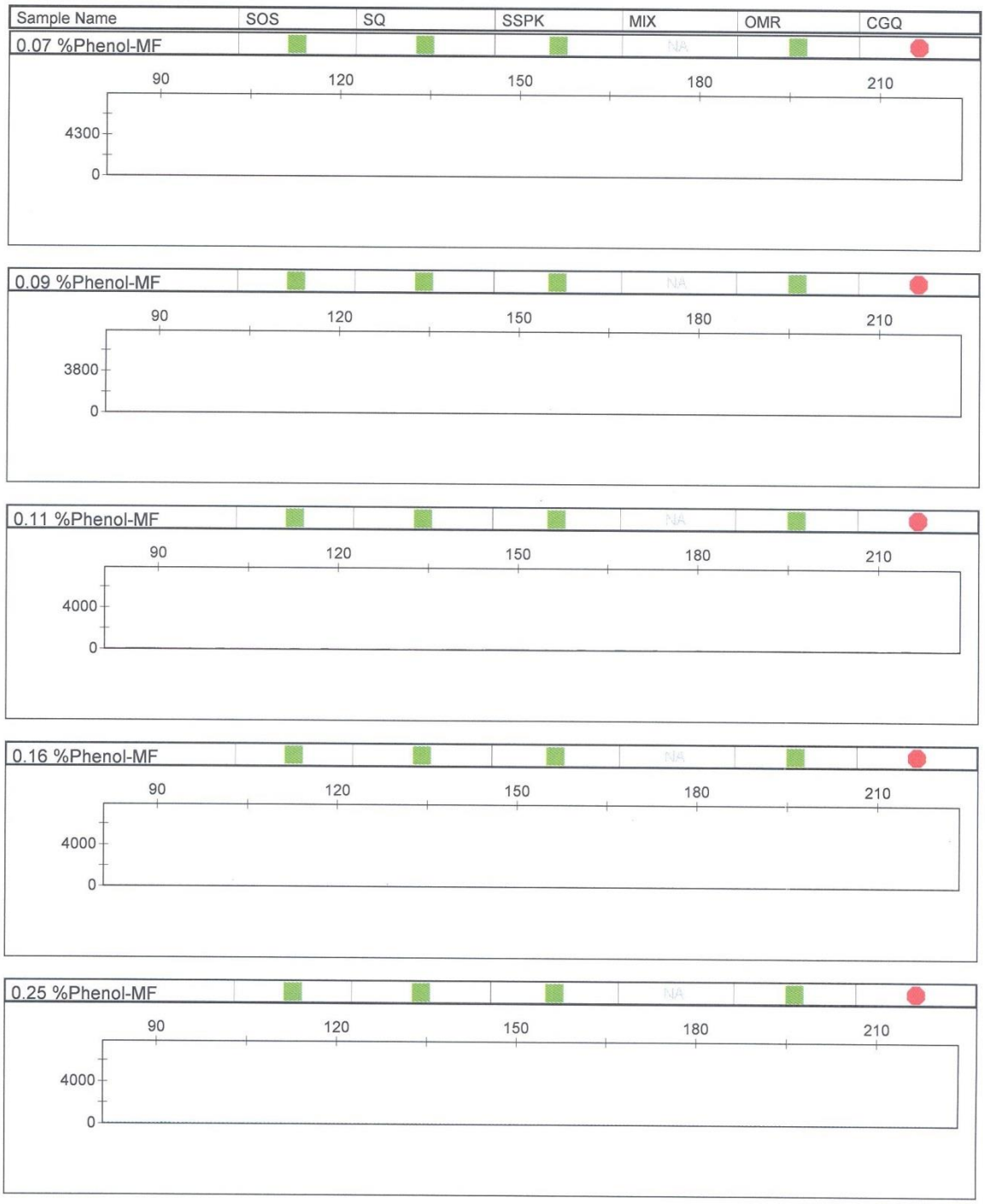

Figure 6.15 Original electropherograms generated from DNA samples (after adding different concentrations of phenol) using Applied Biosystems™ MiniFiler® kit, (n = 3).

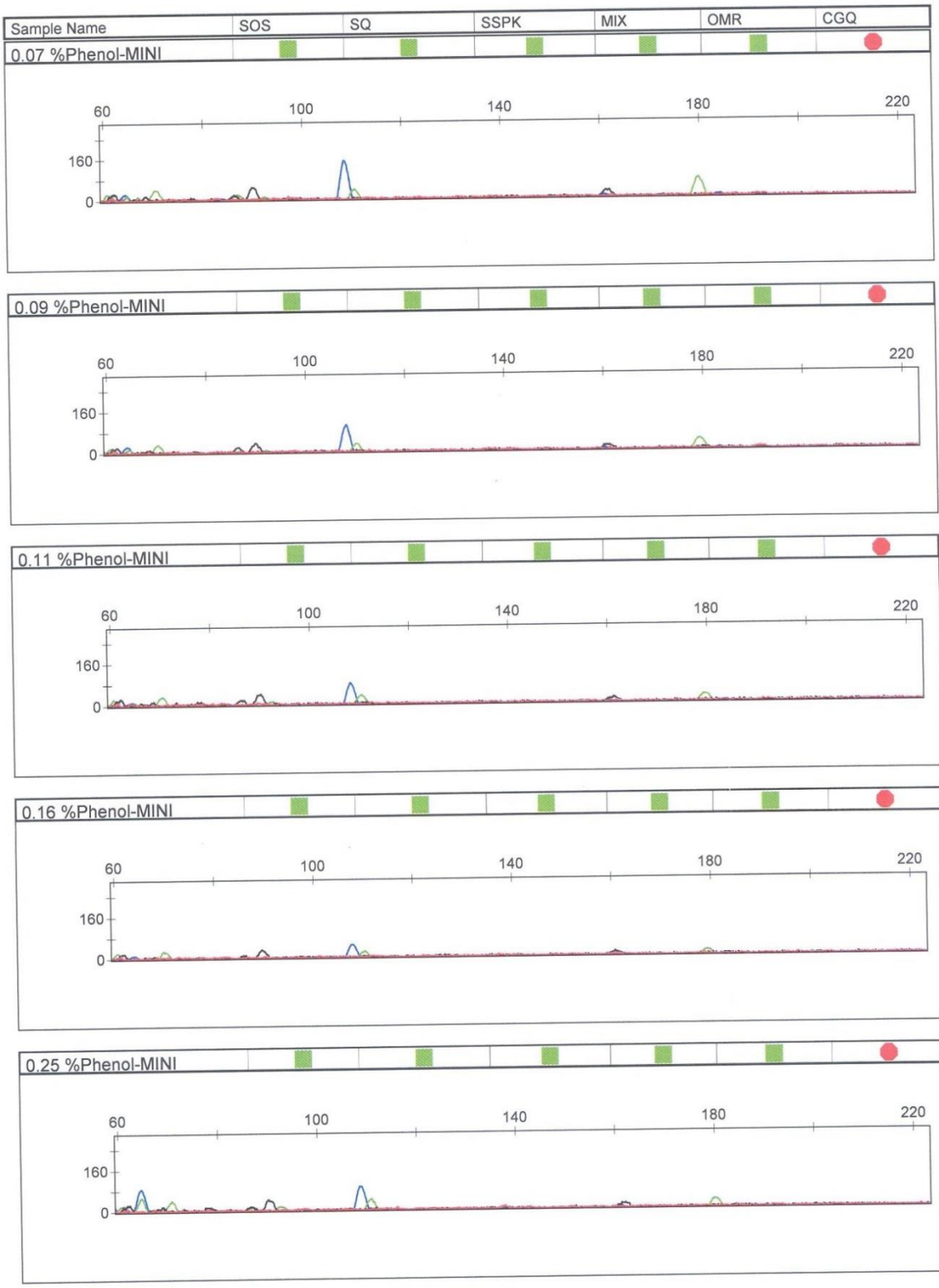

Figure 6.16 Original electropherograms generated from DNA samples (after adding different concentrations of phenol) using mini-INDELs kit, (n = 3).

6.4 DISCUSSION

This study was conducted to compare the amplification efficiency of three PCR assays i.e. Identifiler Plus® (Applied Biosystems™), MiniFiler® (Applied Biosystems™) and mini-INDELs kit to address the issue of degradation and inhibition. Moreover, this study was carried out by designing different experiments keeping in view of real casework scenario.

In the first sequence of experiments, blood-stains were deposited on metal and cloth for three months (June, July and August) in Qatar and were exposed to environmental insults. The temperature was calculated in terms of accumulated degree-days (Larkin *et al.* 2010). It was an attempt to mimic the crime scene situations where blood stains can remain under the sun for long period before they are found by the Police.

The Forensic Scientists often encountered evidence in which biological samples are subjected to cleaning agents. Bleach (NaClO) and hydrogen peroxide (H_2O_2) are the two commonly used cleaning agents used in daily life. Even the presence of their minute quantities in forensic samples might be detrimental to DNA molecule and can lead to its severe degradation (Prince *et al.* 1992; Driessens *et al.* 2009). Second sequence of experiments was performed to determine the effects of bleach and hydrogen peroxide using different concentrations on the DNA samples.

While in the third sequence of experiments, phenol and ethanol were used which are the necessary ingredients of organic method of extracting the DNA (Wallace 1987). Phenol is a part of a mixture of phenol-chloroform-isoamyl alcohol in which components of cell are distributed in three layers after cell lysis from which DNA is suspended in the aqueous layer. Then, this aqueous layer containing DNA is transferred to new tubes and should be washed with washing to remove any residual phenol. If not washed adequately, small traces of phenol might remain in eluted DNA which can inhibit during amplification process. Ethanol is used to precipitate DNA from its aqueous form during extraction process. Ethanol is mostly used in those cases when the volume of DNA elution needs to be concentrated. Before eluting DNA with TE buffer, if not properly dried, it's presence like phenol can inhibit amplification of DNA.

The selection of PCR assays for this study was another important task. Three amplification kits were selected for this purpose and they included Identifiler Plus® (Applied Biosystems™), MiniFiler® (Applied Biosystems™) and mini-INDELs kit (developed during this research). The Identifiler Plus® kit is the widely used PCR assay in the routine forensic laboratories in different parts of the world. It has 15 autosomal STR markers along with amelogenin (size ranging from 75 bp to 450 bp) and it is used for the analysis of case-work (either regular or challenged) and database samples (Identifiler Plus® manual). In contrast, MiniFiler® kit is a complementary kit of Identifiler Plus® kit. It has 8 autosomal STR markers along with amelogenin (size ranging from 75 bp to 230 bp) and it is employed in those cases where high molecular weight amplicons from Identifiler Plus® kit are unable to amplify, so that the lost loci can be possibly recovered (Butler *et al.* 2007). Mini-INDELs kit, which has been developed during this research, is included in this comparative study to evaluate its efficiency and robustness based on its short amplicons (ranging from 75 bp to 130 bp) to genotype the degraded and inhibited DNA samples.

In the first set of degraded samples, where DNA samples were subjected to environmental insults in relation to ADD °C (Larkin *et al.* 2010) for three months, the results indicated that blood samples were degraded completely within three months. The effect of environmental temperature on the quality and quantity of DNA is very negative and amount of DNA available for amplification decreases with increasing temperature (Itani *et al.* 2011). The temperature above 37 °C and 100 % humidity can cause DNA degradation in a short time of 12-16 weeks (Dixon *et al.* 2006).

The amplification results of the samples subjected to environmental insults clearly indicate that mini-INDELs kit leads the other two PCR kits (Identifiler Plus® and MiniFiler®) in terms of completeness of profiles. The DNA samples degraded by environmental insults for 76, 80 and 84 days respectively were unable to generate any profile using Identifiler Plus® kit. While MiniFiler® kit produces profiles for 76 and 80 days, but some drop-outs were observed for sample exposed for 84 days. On the other hand, mini-INDELs were capable of amplifying all of degraded samples which shows its better ability to amplify the degraded DNA samples following environmental insults. The most likely reason for mini-INDELs kit to behave better

than those of STRs is due to its small amplicons (75-130 bp). It was proven in other studies as well that amplification efficiency increases as size of amplicon decreases (Utsuno 2004). On the other hand, both Identifiler Plus® and MiniFiler® kits have larger amplicons (in the ranges of 75-400 bp and 75-220 bp respectively) which resulted in the loss of alleles for higher base pair loci. The results in this study can also be supported by other studies in which it was proven that smaller amplicons of biallelic markers such as SNP and INDELs are amplified better than longer amplicons of STR in case of DNA degradation (Gill *et al.* 1998).

For the second set of degraded samples, the degradation was achieved by adding two oxidizing agents i.e. bleach and hydrogen peroxide in different concentrations. Both MiniFiler® and mini-INDELs kits were found better when amplifying samples treated with bleach and hydrogen peroxide than Identifiler Plus®. Previous studies also indicated that increasing concentrations of these cleaning agents reduces the chances of recovery of DNA profiles (Harris *et al*. 2006; Passi *et al*. 2012).

Regarding the inhibition study, it was ideal to use a number of common inhibitors for the study but due to limited resources, only two inhibitors could be tested. PCR inhibition was introduced to a set of DNA samples using different concentrations of ethanol and phenol. The set of DNA samples contaminated with ethanol were able to amplify only with 5% concentration for all of three kits. Mini-INDELs show weak amplification for higher concentration than 5%, but the results were blow threshold. The amplification efficiency of Identifiler Plus® kit was better than those of MiniFiler® and mini-INDELs while applying on DNA samples contaminated with phenol. It amplifies the samples with 0.07%, 0.09% and 0.11% phenol. On the other hand, MiniFiler® was unable to amplify any of DNA samples mixed with phenol. Mini-INDELs again show weak amplification with low threshold. It is not easy to determine the possible explanation of Identifiler Plus® kit performing better than those of MiniFiler® and mini-INDELs, but possibly the presence of better buffer composition and *Taq* Gold polymerase (Eckhart *et al.* 2000; Identifiler Plus® Manual) and better interaction of polymerase during primer extension (Smith *et al.* 2002) and high melting points of the primers (Chung 2004) might contribute for its better performance.

Inspite of a number of ways to tackle DNA inhibition, little is known regarding their mechanisms. Three possible mechanisms of inhibition have been proposed which include their interaction with DNA template, their binding with polymerases (Bickley *et al.* 1996; Eckhart *et al.* 2000), and their interaction with polymerases during primer extension. At the same time, each type of inhibitor has different way of affecting the amplification process e.g. collagen, calcium and tannic acid inhibit the *Taq* polymerase and compete with magnesium (Bickley *et al.* 1996; Scholz *et al.* 1998), humic acid binds to DNA thus reducing the available sequence of DNA for amplification (Young *et al.* 1993) and hematin and melanin appears to inhibit polymerase activity during primer extension (Akane *et al.* 1994; Larkin *et al.* 1999).

A number of methods have been developed to address the issue of degradation and inhibition during PCR amplification which includes use of bovine serum albumin (Chung 2004), dilution of problematic DNA samples (Opel *et. al.* 2006), use of Microcon filters (Merck Millipore) and silica columns (Yang *et al.* 1998) during extraction process which can help to some extent in retaining larger fragments and eliminating degraded fragments (Noren *et al.* 2013) and use of Centri-Sep columns (ThermoFisher Scientific™). Another approach has been applied by using reduced size PCR products to recover lost loci from degraded and inhibited samples. There are many amplification kits available in the market that can enhance the recovery of DNA profiles from degraded and inhibited DNA samples. STR based MiniFiler® (Applied Biosystems™), PowerPlex® ESI and ESX (Promega™) and INDEL based Investigator® DIPplex (Qiagen™) kits are among those available systems, being used by forensic scientists to process compromised DNA samples.

6.5 CONCLUSION

This study was designed to investigate the efficiency of three different PCR assays to address the issue of DNA degradation and PCR inhibition. Degradation and Inhibition agents were carefully selected keeping in view real case-work scenarios. The results clearly indicated that in most of cases, mini-INDELs kit was found very efficient in comparison to Identifiler Plus® and MiniFiler® kits while applying on the degraded samples. On the other hand, mini-INDELs kit was not very effective for the selected inhibiting agents. A thorough study is needed to check the efficiency of mini-INDELs on commonly encountered PCR inhibitors. In conclusion, it can be said that mini-INDELs kit can be confidently used as an alternative and/or complementary tool while suspecting degradation in DNA samples. In the cases where DNA samples are possibly inhibited, mini-INDELs kit can also be tried to recover the genetic information, but it is recommended to optimise it with common PCR inhibitors, before applying it on the case work inhibited samples.

CHAPTER 7
GENERAL DISCUSSION AND FUTURE WORK

7.1. GENERAL DISCUSSION

Most of the forensic DNA profiling revolves around the short tandem repeats (STRs) in the modern forensic laboratories. The requirement of small quantity of DNA (< 1ng), ability of multiplexing and being polymorphic, make STRs ideal markers for human identification (Schmerer *et al.* 1999; Butler 2003). But STRs have some shortcomings which include their inability to be typed in degraded and inhibited samples, difficulty in mixture interpretation and high mutation rates (Weber *et al.* 2002; Li *et al.* 2011; Manta *et al.* 2012; LaRue *et al.* 2012). Single nucleotide polymorphisms (SNPs) are another type of genetic markers that help forensic scientists in deriving the genetic information from forensic DNA samples. The presence of smaller amplicons, low mutation rates and absence of stutter products are important features, which make them easier to analyze than STRs in degraded DNA (Nachman and Crowell 2000; Brion *et al.* 2005; Dixon *et al.* 2006). But due to less discrimination power, multiple steps procedure and complex interpretation (Philips *et al.* 2007), their use in routine forensic laboratories is limited. Another type of markers named as Insertion Deletion Polymorphisms (INDELs) has been gaining importance by forensic scientists. This is due to their combined advantages of STRs and SNPs, which include the presence of shorter amplicons, low mutation rates, better mixture interpretation and routine genotyping techniques (Romanini *et al.* 2012; Borsting *et al.* 2013; Cereda *et al.* 2014).

7.1.1. Focus of research

This study aimed to estimate the forensic efficiency of autosomal INDELs in Qatar with the possibility of their use as a complementary tool in conjunction with STRs and to evaluate their potential role in the recovery of genetic information from challenged DNA samples (degraded or/and inhibited ones).

In order to achieve these goals, in the first stage, a total of 500 individuals belonging to five different nationalities (Qatari, Pakistani, Sudanese, Tunisian and Yemeni) were genotyped with 30 autosomal INDELs using Investigator DIPplex® kit (Qiagen™). The reason to select Qiagen™ based kit was that a lot of the world's populations with different ethnic origins had already been genotyped with it (Friis *et al.* 2011; Larue *et al.* 2012; Manta *et al.* 2012; Pinto *et al.* 2013; Zaumsegel *et al.* 2013; Ferreira *et al.* 2015; Nunotani *et al.* 2015). Hence it was convenient to compare the forensic efficiency of 30 INDELs data generated during this study to other published results.

7.1.2. Forensic value of INDELs

The statistical data generated from 30 autosomal INDELs revealed that there was no deviation from Hardy-Weinberg Equilibrium and no evidence of linkage disequilibrium for all of 30 markers after Bonferroni's correction. Therefore, the INDEL markers can be considered as independent markers to use for statistical calculations for forensic purposes (Weir *et al.* 1984). The forensic parameters of 30 INDELs for five tested populations were very similar to other populations typed by Investigator® DIPplex kit (Da Silva *et al.* 2013; Wei *et al.* 2013; Pepinski *et al.* 2013; Ferreira *et al* 2015). The Combined Power of Discrimination (CPD) for the 30 INDEL loci was 0.9999999 for all of five populations, which is a satisfactory value for forensic purpose. The Combined Probability of Match (CPM) was calculated in the order of 10^{-12} to 10^{-13} which is higher than 8 mini-STRs in Minifiler® kit (10^{-8} to 10^{-9}), but less than that of 15 STRs of Identifiler Plus® (10^{-15} to 10^{-16}) and 52 SNPs from SNP*for*ID (10^{-17} to 10^{-20}) (Pereira *et al.* 2009).

7.1.3. INDELs as Ancestry Informative Markers

In this study, an attempt was made to derive ancestry information from the genetic data generated during this research. The analysis of INDELs data using Structure software, failed to differentiate the tested populations into five groups. The maximum likelihood calculated by Evanno approach (Evanno *et al*. 2005) was observed, when K = 2. The reason of failure to differentiate populations in this case is due to the use of 30 autosomal INDELs that were selected for identification purpose rather than to derive ancestry information and there was little differentiation of the tested populations. While selecting genetic markers for identification purpose, the criteria are based on the fact that they should have comparable allele frequency

among different populations. On the other hand, ancestry informative markers are selected on the basis of their variable allele frequencies among different population groups (Pfaff *et al.* 2004).

7.1.4. STRs as Ancestry Informative Markers

At the same time, the genetic data of STRs generated during this study, was also attempted to derive ancestry information for five population groups in Qatar. Unlike INDELs, the Structure results for STRs data indicated the presence of four populations i.e. K = 4 which were closer to the number of tested populations i.e. 5. The different outcome of Structure analysis of STRs can be attributed to their polymorphic nature of markers (Gill 2002). The pairwise F_{ST} data showed a close relatedness of Qatari and Yemeni that can be explained by the similar possible similar population origin. Moreover, a Y chromosome study by Cadenas *et al.* (2008) indicated that Qatari and Yemeni populations shared similar Y-haplogroups. Hence, if these two populations are considered as single unit, then the value of K derived from STR data seems quite accurate.

7.1.5. Factors affecting ancestry derivation

In addition to comparable allele frequency pattern of identification INDELs, sample size and self-declared nationality might be two other factors which can influence ancestry derivation. Even it was suggested that 100-150 samples from each population could be enough to calculate profile frequencies (Chakraborty 1992), but it would be better to have larger sample size (sample size of 100 was used for each population in this study) to determine the number of populations and sub-populations accurately. Especially, if population origin of the donor of DNA samples might be different from the declared one, which can result in discontinuities during statistical calculations (Serre *et al.* 2004) and ultimately poor assignment of population. At the same time, the presence of substructures in the main populations might result in wrong deduction of ancestry information. For accurate measurement of ancestry, greater loci or samples would be required (Rosenberg *et al.* 2001).

7.1.6. Multiplexes of INDELs

In addition to Investigator® DIPplex kit (Qiagen™), other INDELs multiplex have also been developed. Some have been developed for forensic identification purposes while others as ancestry informative markers. Recently, a multiplex has been developed with a combination of INDELs and STRs (DIP-STR) that has helped to resolve the unbalanced mixtures efficiently (Oldoni *et al.* 2015; Pinto *et al.* 2015). Following investigator® DIPplex kit, INDELs data have also been generated for 38 INDELs multiplex using different populations (Pereira *et al.* 2009) and they are being used for different forensic applications (Manta *et al.* 2011; Pinto *et al.* 2012; Saiz *et al.* 2014; Martínez-Cortés *et al.* 2015; Ferragut *et al.* 2015; Santarita *et al.* 2015). A number of ancestry based INDEL panels have been developed and applied successfully for the inference of ancestry of different populations (Santos *et al.* 2010; Francez *et al.* 2011; Zaumsegel *et al.* 2012; Li *et al.* 2011; Pereira *et al.* 2012; Romanini *et al.* 2014; Lopes *et al*. 2014). A comparative study was performed by 34-plex SNP test with SNaPshot and 46-plex INDELs. The superiority of 46 INDELs was proved in terms of peak balance of heterozygotes and thus better ability to resolve mixtures (Santos *et al.* 2015). Similar to SNP*for*ID, an online forensic INDELs tool named as *for*InDel (Santos *et al.* 2015), has been created by calculating the allele frequency of 7 major populations including African, European, East Asian, Native American, Oceanian, Middle Eastern, South Asian, using 46 AIM-INDELs (Pereira *et al.* 2012). The genotype of forensic sample of unknown origin can be typed and inference of possible ancestry information will be returned.

7.1.7. Mini-INDELs; A solution for DNA degradation

After proving the forensic efficiency of INDELs to use for DNA case-work, the next stage was set to develop a multiplex PCR assay based on Insertion Deletion polymorphism (INDELs) which can address the issue of DNA degradation. This is encountered commonly in forensic laboratories. To achieve this goal, 14 autosomal INDEL markers were selected and primers were designed for them, keeping in view that they could perform efficiently on degraded DNA samples. The multiplex of 14 INDELs was developed and optimized in a single tube reaction. All the markers were amplified adequately with good peak balance and expected amplicon sizes. The sensitivity of mini-INDELs was found upto 0.03125 ng of genomic DNA with complete and balanced profiles. A concordance study was carried out between mini-INDELs kit and Investigator DIPplex® kit (Qiagen™) for the common loci. The concordance between two kits was observed in 99.7% INDEL alleles. The possible reason of non-concordance of alleles is most likely to point mutation or insertion or/and deletion in the flanking region of primer binding site (Drabek *et al.* 2004; Hill *et al.* 2007).

The efficiency of mini-INDELs PCR assay was evaluated on a set of artificially degraded DNA samples (degradation was achieved by environmental insults and chemical reagents). The successful typing of degraded samples through mini-INDELs kits can be attributed to their smaller PCR products as it was proved that PCR efficiency increases as amplicon size decreases (Utsuno 2004). Also, the outcomes of different studies strengthened the fact that as degradation rate increases, there are more chances of amplifying smaller fragments of DNA than longer ones (Takahashi *et al.* 1997).

The amplification efficiency of mini-INDELs kit was also tested with simulated inhibited DNA samples (induced by ethanol and phenol). Although the possibility of the presence of ethanol and phenol as PCR inhibitors in DNA samples become limited now as new and automated protocols of extracting DNA have been developed and implemented in the forensic laboratories. But some laboratories still choose organic extraction method and in this case, there might be possibility of carry-over of either phenol or ethanol in the eluted DNA. The amplification results of simulated inhibited samples with phenol and ethanol, were not very

encouraging with mini-INDELs kit rather Identifiler Plus kit® (Applied Biosystems™) perform better than mini-INDELs. The reasons of Identifiler Plus kit® amplifying better for phenol/ethanol inhibited samples might be due to the presence of *Taq* Gold polymerase and better buffer composition (Eckhart *et al.* 2000; Identifiler Plus manual), polymerase interaction during primer extension (Smith *et al.* 2002) and high melting points of the primers (Chung *et al.* 2004).

7.1.8. Next Generation Sequencing and forensic DNA profiling

Next generation sequencing (NGS) provides a latest genetic tool for forensic DNA analysis. It is now possible to analyze a combination of different genetic markers (autosomal STRs, Y-STRs, SNPs, INDELs, mtDNA) in a single experiment, which cannot be done by using standard PCR and CE systems (Børsting and Morling 2015). Although most of NGS work has been done on SNPs, but STRs being integral part of regular forensic DNA analysis, cannot be ignored. STR typing using next generation sequencing also helped in different scenario of forensic casework. While next generation sequencing of STRs, it was observed that about 30% of the homozygous allele calls by capillary electrophoresis turned out to be heterozygous (Rockenbauer *et al.* 2014). The sequencing of STRs by NGS technology helped in determining exact number of repeats and ultimately makes it simpler to interpret mixtures. It becomes easier now to differentiate between real minor contributor peaks and stutter peaks of STR mixtures through NGS (Dalsgaard *et al.* 2013).

Applied Biosystems™ offers two commercial NGS kits (Identity panel and Ancestry panel) based on SNPs by using Ion torrent™ PGM system. At the same time, a panel of core STRs named as HID STR 10-plex (based on loci which are less than 170 base pairs) has been released for initial testing. In its preliminary results, it has been proved successful in typing degraded samples (Fordyce *et al.* 2014). On the other hand, Illumina® has launched its multiplex NGS kit called as Forenseq™ DNA signature which can amplify 27 autosomal STRs, 8 X-STRs, 25 Y-STRs, 95 autosomal SNPs for human identification, 56 autosomal SNPs for ancestry and 24 autosomal SNPs for pigmentary traits. Forenseq™ kit can be used on Illumina® MiSeq FGx platform. Due to its recent launch, Forenseq™ kit is still under testing in different labs.

One important feature of NGS platform is the possibility of making custom based panel like available commercial panels. Forensic scientists or organizations can make their own custom panel for any combination of genetic markers and can use either of platform (Illumina® MiSeq FGx platform or Applied Biosystems™ Ion torrent™ PGM system) to analyze the data in shorter time and cost effectively. As a lot of research is being done on INDELs, it is now possible to create an INDEL panel containing as many as markers (identity or/and AIMs), to generate huge amount of population data using NGS platform.

The major disadvantage of using NGS technology as a regular genetic tool in a routine forensic laboratory is due to their longer procedures, which take 2-3 days. There is also a need of developing the sophisticated software (to analyze NGS data conveniently) and a proper system of reporting (Treangen *et al.* 2012).

7.2. Conclusions

After a thorough statistical analysis of INDELs, It can be concluded that they are sufficiently polymorphic and can be used for forensic purposes. The possible forensic applications where INDELs can be applied include genotyping for degraded and low level DNA samples, mixture interpretation and deficient paternity cases. These applications can be attributed due to their shorter amplicons, absence of stutter products and low mutation rates.

Autosomal INDELs contained in Investigator® DIPplex kit (Qiagen™) are not ideal markers for population differentiation due to their comparable allele frequencies present in different populations.

The mini-INDELs kit developed during this study, provides an efficient solution for the recovery of genetic information from challenged DNA samples. The validation studies performed on mini-INDELs during this study showed that they are very effective in deriving the lost genetic information from compromised DNA samples. Hence, they can be employed confidently in conjunction with STRs as a complementary tool in regular forensic case-work, especially in

those cases where low level of DNA (such as touch DNA) and DNA degradation (such as in mass disaster cases) are suspected.

As a conclusion, it can be said confidently that INDELs in general and mini-INDELs in particular can be used as useful forensic tools to help in solving forensic DNA casework.

7.3. Limitations

During this research, a successful effort was made to calculate forensic value of INDELs by using five populations based in Qatar. But if more populations could be included in this study, forensic efficiency of INDELs can be evaluated in a better way.

Although, Mini-INDELs kit contains 14 markers along with sex discriminating amelogenin and have sufficient polymorphic value but still more markers could be accommodated to increase its forensic value.

During this study, a comprehensive study was performed to evaluate the efficiency of mini-INDELs to type challenged samples. If more variety of challenged samples (such as different types of inhibited samples) could be included then better evaluation of mini-INDELs kit could be done.

During concordance study between Investigator® DIPplex kit and mini-INDELs kit, a total of 10 non-concordant alleles were detected. It would be ideal to investigate the exact reason of non-concordance of these alleles by sequencing them.

7.4. Recommendations

The climate of the GCC States (Saudi Arabia, United Arab Emirates, Kuwait, Oman, Bahrain and Qatar), is like tropical desert and especially in summers when the weather is extremely hot and humid, temperature reach upto 50 °C. DNA samples exposed to longer period at this high temperature on the crime scenes might result in severe DNA degradation. It would be a challenging task to recover the genetic information lost from these challenged samples. There should be some protocols, which need to be implemented properly in order to recover as much

as possible genetic information from these challenged samples. Starting from crimes scene collection, it is necessary that crime scene officers are trained adequately to understand the processes of sample collection, preservation and transportation, so that the quality of collected forensic evidence will not be compromised. Initial screening and sampling procedures of evidences during examination of forensic evidences inside laboratory should be performed carefully making the possibility of any type of contamination minimum. Suspecting DNA degradation or/and inhibition, proper and optimized extraction method should be chosen. Any modification in the normal standard operational procedures should not be made without prior validation. The use of additional tools like Microcon, silica column or Centricep during extraction of DNA can be made depending on the quality of sample. Use of Quantifiler® human Trio (Applied Biosystems™) or any other available quantification tool can be used which can help in assessing the quality and quantity of DNA based on degradation index or/and internal positive control (IPC) values. After quantification, DNA quantities can be normalized depending on the results and PCR should be performed by using routine STR chemistry. If the PCR result of STRs shows the signs of either degradation or inhibition such as differential or preferential amplification or no amplification, then supplementary amplification tools such as mini-STRs and other available tools can be tried to recover the lost loci. Mini-INDELs kit developed in this study can be used as an alternative tool to type the degraded samples.

7.5. FUTURE WORK

According to Qatar statistics (2015), a total of 63 nationalities are currently residing in Qatar including Indian (24 %), Nepalese (17 %), Qatari (12 %), Pilipino (9%), Egyptians (8 %), Bangladeshis (7%), Pakistanis (2.5%), Sri Lankans (1.5%) and many others. In order to use INDELs in the regular forensic case work stream like STRs, more populations based in Qatar, can be included in the future studies and their forensic efficiency and other statistical parameters can be evaluated.

Mini-INDELs kit can be extended to accommodate more INDEL markers in its multiplex. Additional INDEL markers can improve the forensic efficiency of the existing ones.

During this study, an effort was made to evaluate the PCR efficiency of mini-INDELs to type degraded and inhibited DNA samples at a limited scale. In order to get better evaluation of mini-INDELs kit to address the issues of DNA degradation/inhibition, different types of degrading and inhibiting DNA samples can be tested which can give better understanding about the types of forensic challenged samples which can be successfully typed with mini-INDELs.

Different Ancestry informative INDELs multiplexes have been developed and applied to some populations to estimate their ethnic origins (Pablo *et al.* 2011; Zaumsegel *et al.* 2012; Pereira *et al.* 2012). These AIMS-INDELs can be applied to the known populations' samples available in Qatar to estimate their ancestry information.

Another emerging tool, Next generation sequencing (NGS) can possibly be used to analyse INDELs (for the purposes of identification and derivation of ancestry information) by using a large number of samples conveniently and cost-effectively. At the same time, NGS can help in the investigation of the situations including non-concordance between alleles generated from mini-INDELs and investigator® DIPplex kits during this study where INDELs are needed to be sequenced.

REFERENCES

AKANE, A., MATSUBARA, K., NAKAMURA, H., TAKAHASHI, S. & KIMURA, K (1994) Identification of the Heme Compound Copurified with Deoxyribonucleic-Acid (DNA) from Bloodstains, a Major Inhibitor of Polymerase Chain-Reaction (PCR) Amplification. *Journal of Forensic Sciences,* 39(2): 362-372.

AKHTERUZZAMAN, S., DAS, S.A., HOSEN, I. and FERDOUS, A (2013) Genetic polymorphism of 30 InDel markers for forensic use in Bangladeshi population. *Forensic Science International: Genetics Supplement Series,* 4(1): 348-349.

ALAEDDINI, R., WALSH, S.J. and ABBAS, A (2010) Forensic implications of genetic analyses from degraded DNA-A review. *Forensic Science International-Genetics*, 4(3): 148-157.

ALENIZI, M.A., GOODWIN, W., HADI, S., ALENIZI, H.H., ALTAMAR, K.A. and ALSIKEL, M.S (2009) Concordance between the AmpFlSTR® MiniFiler™ and AmpFlSTR® Identifiler® PCR amplification kits in the Kuwaiti population. *Journal of Forensic Sciences,* 54(2): 350-352.

AL-SOUD, W. & RADSTROM, P (1998) Capacity of nine thermostable DNA polymerases to mediate DNA amplification in the presence of PCR-Inhibiting samples. *Applied and Environmental Microbiology,* 64(10): 3748-3753

AL-SOUD, W.A. and RÅDSTRÖM, P (2001) Purification and characterization of PCR-inhibitory components in blood cells. *Journal of Clinical Microbiology,* 39(2): 485-493.

ALTSHULER, D.M., GIBBS, R.A., PELTONEN, L., DERMITZAKIS, E., SCHAFFNER, S.F., YU, F., BONNEN, P.E., DE BAKKER, P.I.W., DELOUKAS, P., GABRIEL, S.B and INTERNATIONAL HapMap 3 Consortium (2010) Integrating common and rare genetic variation in diverse human populations*. Nature*, 467(7311): 52-58.

AMORIM, A. and PEREIRA, L (2005) Pros and cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs. *Forensic Science International,* 150(1): 17-21.

ASARI, M., OKUDA, K., HOSHINA, C., OMURA, T., TASAKI, Y., SHIONO, H., MATSUBARA, K. and SHIMIZU, K (2015) Multicolor-based discrimination of 21 short tandem repeats and amelogenin using four fluorescent universal primers. *Analytical Biochemistry,* 49(4): 16-22.

ASARI, M., OKUDA, K., YAJIMA, D., MASEDA, C., HOSHINA, C., OMURA, T., SHIONO, H., MATSUBARA, K. and SHIMIZU, K (2015) Rapid genotyping of 25 autosomal STRs in a Japanese population using fluorescent universal primers containing locked nucleic acids. *Journal of Forensic and Legal Medicine,* 31(3): 36-41.

ASARI, M., UMETSU, K., ADACHI, N., AZUMI, J., SHIMIZU, K. and SHIONO, H (2007) Utility of haplogroup determination for forensic mtDNA analysis in the Japanese population. *Journal of Legal Medicine,* 9(5): 237-240.

BAECHTEL, F.S., SMERICK, J.B., PRESLEY, K.W. and BUDOWLE, B (1993) Multigenerational amplification of a reference ladder for alleles at locus D1S80. *Journal of Forensic Sciences,* 38(5): 1176-1182.

BALLANTYNE, K.N., KEERL, V., WOLLSTEIN, A., CHOI, Y., ZUNIGA, S.B., RALF, A., VERMEULEN, M., DE KNIJFF, P. and KAYSER, M (2012) A new future of forensic Y-chromosome analysis: Rapidly mutating Y-STRs for differentiating male relatives and paternal lineages. *Forensic Science International: Genetics,* 6(2): 208-218.

BALLARD, L.W., ADAMS, P.S., BAO, Y., BARTLEY, D., BINTZLER, D., KASCH, L., PETUKHOVA, L. and ROSATO, C (2002) Strategies for genotyping: Effectiveness of tailing primers to increase accuracy in short tandem repeat determinations. *Journal of Biomolecular Techniques,* 13(1): 20-29.

BARBARO, A., CORMACI, P. and BARBARO, A (2006a) DNA typing from 15-year-old bloodstains. *International Congress Series,* 1288: 550-552.

BARBARO, A., CORMACI, P. and BARBARO, A (2006b) LCN DNA typing from touched objects. *International Congress Series,* 1288: 553-555.

BARBARO, A., CORMACI, P. and BARBARO, A (2006c) Multiplex STRs amplification from hair shaft: Validation study. *International Congress Series,* 1288: 676-678.

BARBARO, A., CORMACI, P. and BARBARO, A (2006d) X-STR typing for an identification casework. *International Congress Series,* 1288: 513-515.

BARBUJANI, G., & COLONNA, V (2010) Human genome diversity: frequently asked questions. *Trends in Genetics*, 26(7): 285-295.

BARNHOLTZ-SLOAN, J.S., PFAFF, C.L., CHAKRABORTY, R. and LONG, J.C (2005) Informativeness of the CODIS STR loci for admixture analysis. *Journal of Forensic Sciences,* 50(6): 1322-1326.

BASTOS-RODRIGUES, L., PIMENTA, J.R. and PENA, S.D.J (2006) The genetic structure of human populations studied through short insertion-deletion polymorphisms. *Annals of Human Genetics,* 70(5): 658-665.

BECK, C.R., COLLIER, P., MACFARLANE, C., MALIG, M., KIDD, J.M., EICHLER, E.E., BADGE, R.M. and MORAN, J.V (2010) LINE-1 retrotransposition activity in human genomes. *Cell,* 141(7): 1159-1170.

BENDER, K., FARFAN, M. & SCHNEIDER, P (2004) Preparation of degraded human DNA under controlled conditions. *Forensic Science International,* 139(3): 135-140.

BENSON, G. and DONG, L (1999) Reconstructing the duplication history of a tandem repeat. *ISMB. International Conference on Intelligent Systems for Molecular Biology, 412(2):* 44-53.

BENNETT, E.A., COLEMAN, L.E., TSUI, C., PITTARD, W.S. and DEVINE, S.E (2004) Natural genetic variation caused by transposable elements in humans. *Genetics,* 168(2): 933-951.

BHANGALE, T.R., RIEDER, M.J., LIVINGSTON, R.J. and NICKERSON, D.A (2005) Comprehensive identification and characterization of diallelic insertion-deletion polymorphisms in 330 human candidate genes. *Human Molecular Genetics,* 14(1): 59-69.

BHANGALE, T.R., STEPHENS, M. and NICKERSON, D.A (2006) Automating re-sequencing-based detection of insertion-deletion polymorphisms. *Nature Genetics,* 38(12): 1457-1462.

BICKLEY J, SHORT JK, MCDOWELL G, PARKES HC (1996) Polymerase chain reaction (PCR) detection of *Listeria monocytogenes* in diluted milk and reversal of PCR inhibiton caused by calcium ions. Applied Microbiology, *22(8)*: 153-155.

BORSTING, C. and MORLING, N (2011) Mutations and/or close relatives? Six case work examples where 49 autosomal SNPs were used as supplementary markers*. Forensic Science International-Genetics*, 5(3): 236-241.

BORSTING, C., MOGENSEN, H.S. and MORLING, N (2013) Forensic genetic SNP typing of low template DNA and highly degraded DNA from crime case samples*. Forensic Science International: Genetics,* 5(7): 345-352.

BØRSTING, C. and MORLING, N (2015) Next generation sequencing and its applications in forensic genetics. *Forensic Science International: Genetics,* 18(3): 78-89.

BOURKE MT, SCHERCZINGER CA, LADD C, LEE HC (1999) NaOH treatment to neutralize inhibitors of *Taq* polymerase. *Journal of Forensic Science,* 44 (5): 1046–50.

BOWCOCK, A. M., RUIZ-LINARES, A., TOMFOHRDE, J., MINCH, E., KIDD, J. R., & CAVALLI-SFORZA, L. L (1994) High resolution of human evolutionary trees with polymorphic microsatellites. *Nature*, 368 (6470): 455-457.

BRAID, M.D., DANIELS, L.M. and KITTS, C.L (2003) Removal of PCR inhibitors from soil DNA by chemical flocculation. *Journal of Microbiological Methods,* 52(3): 389-393.

BRANDSTÄTTER, A., SALAS, A., NIEDERSTÄTTER, H., GASSNER, C., CARRACEDO, A. and PARSON, W (2006) Dissection of mitochondrial superhaplogroup H using coding region SNPs. *Electrophoresis,* 27(13): 2541-2550.

BRIÓN, M., SANCHEZ, J., BALOGH, K., THACKER, C., BLANCO-VEREA, A., BØRSTING, C., STRADMANN-BELLINGHAUSEN, B., BOGUS, M., SYNDERCOMBE-COURT, D., SCHNEIDER, P., CARRACEDO, A. and MORLING, N (2005) Introduction of an single nucleotide polymorphism based? Major Y-chromosome haplogroup typing kit? suitable for predicting the geographical origin of male lineages*. Electrophoresis,* 26(23): 4411-4420.

BROMAN, K.W., MURRAY, J.C., SHEFFIELD, V.C., WHITE, R.L. and WEBER, J.L (1998) Comprehensive human genetic maps: Individual and sex-specific variation in recombination. *American Journal of Human Genetics,* 63(3): 861-869.

BROMAN, K.W. and WEBER, J.L (1998) Estimation of pairwise relationships in the presence of genotyping errors. *American Journal of Human Genetics,* 63(5): 1563-1564.

BUDIMLIJA, Z.M., PRINZ, M.K., ZELSON-MUNDORFF, A., WIERSEMA, J., BARTELINK, E., MACKINNON, G., NAZZARUOLO, B.L., ESTACIO, S.M., HENNESSEY, M.J. and SHALER, R.C (2003) World Trade Center human identification project: Experiences with individual body identification cases. *Croatian Medical Journal,* 44(3): 259-263.

BUDOWLE B, ET AL (1985) Population data on the forensic genetic markers: phosphoglucomatase-1, esterase D, erythrocyte acid phosphatase and glyoxylase I. *Forensic Science International,* 45(8):77-81.

BUDOWLE B., MASIBAY A., ANDERSON S.J., BARNA C., BIEGA L., BRENNEKE S., BROWN B.L (2001) STR primer concordance study*. Forensic Science International*, 124(4): 47-54.

BUSE, E.L., PUTINIER, J.C., HONG, M.M., YAP, A.E. and HARTMANN, J.M (2003) Performance evaluation of two multiplexes used in fluorescent short tandem repeat DNA analysis. *Journal of Forensic Sciences,* 48(2): 348-357.

BUTLER, J. M (2003) The development of reduced size STR amplicons as tools for analysis of degraded DNA. *Journal of Forensic Sciences, 48*(5): 1054-1064 .

BUTLER, J.M., BISHOP, D.T. and BARRETT, J.H (2005) Strategies for selecting subsets of single-nucleotide polymorphisms to genotype in association studies. *BMC Genetics,* 6(1): 154-155.

BUTLER, J.M (2005) Forensic DNA Typing: Biology, Technology and Genetics for STR Markers. Second Edition ed. Burlington, MA, USA: *Elsevier Academic Press*, 5(12): 452-458.

BUTLER, J. M (2006) Genetics and genomics of core short tandem repeat loci used in human identity testing. *Journal of Forensic Sciences, 51(3):* 253-265.

BUTLER, J. M (2007) Short tandem repeat typing technologies used in human identity testing. *BioTechniques, 43*(4): 144-146.

BUTLER, J.M (2010) Forensic Challenges: Degraded DNA, Mixtures, and LCN. *Fundamentals of Forensic DNA Typing, 55(4):* 315-339.

BUTLER, J.M (2012) Short tandem repeats (STR) loci and kits. *Advanced Topics in Forensic DNA Typing,* 44(5): 293-309

CADENAS, A.M., ZHIVOTOVSKY, L.A., CAVALLI-SFORZA, L.L., UNDERHILL, P.A. and HERRERA, R.J (2008) Y-chromosome diversity characterizes the Gulf of Oman. *European Journal of Human Genetics,* 16(3): 374-386.

CANN, H.M., DE TOMA, C., CAZES, L., LEGRAND, M.F., MOREL, V., PIOUFFRE, L., BODMER, J.,BODMER, W.F., BONNE-TAMIR, B., CAMBON-THOMSEN, A (2002) A human genome diversity cell line panel. *Science*, 296(5566): 261-262.

CARDENA, M.M.S.G., RIBEIRO-DOS-SANTOS, Â., SANTOS, S., MANSUR, A.J., PEREIRA, A.C. and FRIDMAN, C (2013) Assessment of the Relationship between Self-Declared Ethnicity, Mitochondrial Haplogroups and Genomic Ancestry in Brazilian Individuals. *PLoS ONE,* 8(4): 245-247.

CAROL R. E. and MELVIN E (2001) *Countries and Their Cultures*. Macmillan Reference USA, 44(3): 18-25.

CARVALHO, A., CAINÉ, L., CARVALHO, R. and PINHEIRO, M.F (2011) Application of indels (investigator DIPplex) in mixture samples. *Forensic Science International: Genetics Supplement Series,* 3(1): 351-352.

CAVALLI-SFORZA, L.L (2005) Opinion - The Human Genome Diversity Project: past, present and future. *Nature Reviews Genetics*, 6(4): 333-340.

CEREDA, G., BIEDERMANN, A., HALL, D. and TARONI, F (2014) Object-oriented Bayesian networks for evaluating DIP–STR profiling results from unbalanced DNA mixtures. *Forensic Science International: Genetics,* 8(1): 159-169.

CHAKRABORTY, R (1992) Sample size requirements for addressing the population genetic issues of forensic use of DNA typing. *Human Biology*, 64(2): 141-159.

CHAKRABORTY, R., STIVERS, D.N., SU, B., ZHONG, Y.X. and BUDOWLE, B (1999) The utility of short tandem repeat loci beyond human identification: Implications for development of new DNA typing systems. *Electrophoresis,* 20(8): 1682-1696.

CHEN, K., MCLELLAN, M.D., DING, L., WENDL, M.C., KASAI, Y., WILSON, R.K. and MARDIS, E.R (2007) PolyScan: An automatic indel and SNP detection approach to the analysis of human re-sequencing data. *Genome Research,* 17(5): 659-666.

CHUNG, D.T., DRÁBEK, J., OPEL, K.L., BUTLER, J.M. and MCCORD, B.R (2004) A study on the effects of degradation and template concentration on the amplification efficiency of the STR Miniplex primer sets. *Journal of Forensic Sciences,* 49(4): 733-740.

CLAYTON TM, HILL SM, DENTON LA, WATSON SK, URQUHART A.J (2004) Primer binding site mutations affecting the typing of STR loci contained within the AMPFlSTR SGM Plus kit. *Forensic Science International,* 55(3): 39-44.

COBLE, M.D. & BUTLER, J. M (2005) Characterization of new MiniSTR Loci aid analysis of degraded DNA. *Journal of Forensic Science International,* 50(1): 43-53.

COLLINS P.J., H. L., LEIBELT C.S., ROBY R.K., REEDER D.J., FOXALL P.A (2003) Developmental validation of a single-tube amplification of the 13 CODIS STR loci, D2S1338, D19S433, and amelogenin: the AmpF/STR® Identifiler PCR Amplification Kit. *Journal of Forensic Science,* 49(4): 1265-1277.

COMAS, D., CALAFELL, F., BENCHEMSI, N., HELAL, A., LEFRANC, G., STONEKING, M., BATZER, M.A., BERTRANPETIT, J. and SAJANTILA, A (2000) *Alu* insertion polymorphisms in NW Africa and the Iberian Peninsula: Evidence for a strong genetic boundary through the Gibraltar Straits. *Human genetics,* 107(4): 312-319.

COMEY, C., KOONS, B., PRESLEY, K., SMERICK, J., SOBIERALSKI, C., STANLEY, D. & BAECHTEL, F (1994) DNA Extraction Strategies for Amplified Fragment Length Polymorphism Analysis. *Journal of Forensic Sciences,* 39(5): 1254-1269.

CURTIN, F. and SCHULZ, P (1998) Multiple correlations and Bonferroni's correction. *Biological psychiatry,* 44(8): 775-777.

D' AMATO, M.E., HEFKE, G. and DAVISON, S (2015) Evaluation of DIPplex investigator kit in European, Asian and African populations. *Forensic Science International: Genetics, 44(*5): 470-471.

DALSGAARD, S., ROCKENBAUER, E., GELARDI, C., BØRSTING, C., FORDYCE, S.L. and MORLING, N (2013) Characterization of mutations and sequence variations in complex STR loci by second generation sequencing. *Forensic Science International: Genetics,* 4(1): 218-219.

DA SILVA, C.V., MATOS, S., COSTA, H.A., MORAIS, P., DOS SANTOS, R.M., ESPINHEIRA, R., SANTOS, J.C. and AMORIM, A (2013) Genetic portrait of south Portugal population with InDel markers. *Forensic Science International: Genetics,* 7(4): 101-103.

DAWSON, E., CHEN, Y., HUNT, S., SMINK, L.J., HUNT, A., RICE, K., LIVINGSTON, S., BUMPSTEAD, S., BRUSKIEWICH, R., SHAM, P., GANSKE, R., ADAMS, M., KAWASAKI, K., SHIMIZU, N., MINOSHIMA, S., ROE, B., BENTLEY, D. and DUNHAM, I (2001) A SNP resource for human chromosome 22: Extracting dense clusters of SNPs from the genomic sequence. *Genome research,* 11(1): 170-178.

DIXON, L.A., DOBBINS, A.E., PULKER, H.K., BUTLER, J.M., VALLONE, P.M., COBLE, M.D., PARSON, W., BERGER, B., GRUBWIESER, P., MOGENSEN, H.S., MORLING, N., NIELSEN, K., SANCHEZ, J.J., PETKOVSKI, E., CARRACEDO, A., SANCHEZ-DIZ, P., RAMOS-LUIS, E., BRION, M., IRWIN, J.A., JUST, R.S., LOREILLE, O., PARSONS, T.J., SYNDERCOMBE-COURT, D., SCHMITTER, H., STRADMANN-BELLINGHAUSEN, B., BENDER, K. and GILL, P (2006) Analysis of artificially degraded DNA using STRs and SNPs-results of a collaborative European (EDNAP) exercise. *Forensic Science International,* 164(1): 33-44.

DRABEK J, CHUNG DT, BUTLER JM, MCCORD B.R (2004) Concordance Study between Miniplex Assays and a Commercial STR Typing Kit. *Journal of Forensic Science*, 48(4): 859-860.

DRIESSENS, N., VERSTEYHE, S., GHADDHAB, C., BURNIAT, A., DE DEKEN, X., VAN SANDE, J., DUMONT, J., MIOT, F. and CORVILAIN, B (2009) Hydrogen peroxide induces DNA single- and double-strand breaks in thyroid cells and is therefore a potential mutagen for this organ. *Endocrine-Related Cancer,* 16(3): 845-856.

ECKHART L, BACH J, BAN J, TSCHACHLER E (2000) Melanin binds reversibly to thermostable DNA polymerase and inhibits its activity. *Biochemical Biophysical Research Communication*, 271(4): 726-730.

ELLEGREN, H (2000) Microsatellite mutations in the germline: Implications for evolutionary inference. *Trends in Genetics,* 16(12): 551-558.

EVANNO, G., REGNAUT, S. and GOUDET, J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology,* 14(8): 2611-2620.

EXCOFFIER, L., HOFER, T. and FOLL, M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity,* 103(4): 285-298.

EXCOFFIER, L. and LISCHER, H.E.L (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources,* 10(3): 564-567.

FERRAGUT, J.F., PEREIRA, R., CASTRO, J.A., RAMON, C., NOGUEIRO, I., AMORIM, A. and PICORNELL, A (2016) Genetic diversity of 38 insertion–deletion polymorphisms in Jewish populations. *Forensic Science International: Genetics,* 21(4): 1-4.

FERREIRA PALHA,TERESINHA DE JESUS BRABO, RIBEIRO RODRIGUES, E.M., CAVALCANTE, G.C., MARRERO, A., DE SOUZA, I.R., SEKI UEHARA, C.J., SILVEIRA DA MOTTA,CARLOS HENRIQUE ARES, KOSHIKENE, D., DA SILVA, D.A., DE CARVALHO, E.F., CHEMALE, G., FREITAS, J.M., ALEXANDRE, L., PARANAIBA, R.T.F., SOLER, M.P. and SANTOS, S (2015) Population genetic analysis of insertion–deletion polymorphisms in a Brazilian population using the Investigator DIPplex kit. *Forensic Science International: Genetics,* 19(1): 10-14.

FONDEVILA, M., PHILLIPS, C., NAVERAN, N., FERNANDEZ, L., CEREZO, M., SALAS, A., CARRACEDO, A. and LAREU, M.V (2008) Case report: Identification of skeletal remains using short-amplicon marker analysis of severely degraded DNA extracted from a decomposed and charred femur. *Forensic Science International: Genetics,* 2(3): 212-218.

FONDEVILA, M., PEREIRA, R., GUSMÃO, L., PHILLIPS, C., LAREU, M.V., CARRACEDO, A., BUTLER, J.M. and VALLONE, P.M (2011) Forensic performance of insertion-deletion marker systems. *Forensic Science International: Genetics,* 3(1): 443-444.

FONDEVILA, M., PHILLIPS, C., SANTOS, C., FREIRE ARADAS, A., VALLONE, P.M., BUTLER, J.M., LAREU, M.V. and CARRACEDO, Á (2013) Revision of the SNPforID 34-plex forensic ancestry test: Assay enhancements, standard reference sample genotypes and extended population studies. *Forensic Science International: Genetics,* 7(1): 63-74.

FORDYCE, S.L., MOGENSEN, H.S., BØRSTING, C., LAGACÉ, R.E., CHANG, C.-., RAJAGOPALAN, N. and MORLING, N (2014) Second-generation sequencing of forensic STRs using the Ion Torrent™ HID STR 10-plex and the Ion PGM™. *Forensic Science International: Genetics,* 14(1): 132-140.

FRANCEZ, P.A.D.C., RIBEIRO-RODRIGUES, E.M. and DOS SANTOS, S.E.B (2012) Allelic frequencies and statistical data obtained from 48 AIM INDEL loci in an admixed population from the Brazilian Amazon. *Forensic Science International: Genetics,* 6(1): 132-135.

FRAZER, K.A., BALLINGER, D.G., COX, D.R., HINDS, D.A., STUVE, L.L., GIBBS, R.A., BELMONT, J.W., BOUDREAU, A., HARDENBOL, P., ET AL. AND INTERNATIONAL HAPMAP CONSORTIUM (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature,* 449(7164): 0028-0836.

FRIIS, S.L., BØRSTING, C., ROCKENBAUER, E., POULSEN, L., FREDSLUND, S.F., TOMAS, C. and MORLING, N (2011) Typing of 30 insertion/deletions in Danes using the first commercial indel kit - Mentype ® DIPplex. *Forensic Science International: Genetics,* 6(2): 72-74.

FRUDAKIS, T. VENKATESWARLU, M.J. THOMAS, Z. GASKIN, S. GINJUPALLI, S. GUNTURI, V. PONNUSWAMY, S. NATARAJAN, P.K. NACHIMUTHU (2003) A classifier for the SNP-based inference of ancestry, *Journal of Forensic Science,* 48(2): 771–782.

FRUDAKIS, T. HALDER, M. SHRIVER, M. THOMAS, J.R. FERNANDEZ, T (2008) A panel of ancestry informative markers for estimating individual bio-geographicalal ancestry and admixture from four continents: utility and applications. *Human Mutation*, 29(3): 648–658.

GILL, P., D'ALOJA, E., DUPUY, B., ERIKSEN, B., JANGBLAD, M., JOHNSSON, V., KLOOSTERMAN, A.D., KRATZER, A., LAREU, M.V., MEVAG, B., MORLING, N., PHILLIPS, C., PFITZINGER, H., RAND, S., SABATIER, M., SCHEITHAUER, R., SCHMITTER, H., SCHNEIDER, P., SKITSA, I. and VIDE, M.C (1998) Report of the European DNA profiling group (EDNAP)-an investigation of the hypervariable STR loci ACTBP2, APOAI1 and D11S554 and the compound loci D12S391 and D1S1656. *Forensic Science International,* 98(3): 193-200.

GILL P, WHITAKER J, FLAXMAN C, BROWN N, BUCKLETON J (2001) *An investigation of the rigor of interpretation rules of STRs derived from less than 100 pg of DNA. Forensic Science International*, 112(4): 17-40.

GOLENBERG, E.M., CLEGG, M.T., DURBIN, M.L., DOEBLEY, J. and MA, D.P (1993) Evolution of a Noncoding Region of the Chloroplast Genome. *Molecular Phylogenetics and Evolution,* 2(1): 52-64.

GOLENBERG, E. M., BICKEL, A. & WEIHS, P (1996) Effect of highly fragmented DNA on PCR. *Nucleic Acids Research,* 24(24): 5026-5033.

GRIFFITHS, R. A., ET AL (1998). New reference allelic ladders to improve allelic designation in a multiplex STR system. *International Journal of Legal Medicine, 111(3):* 267 – 272.

GRIMES, E.A., NOAKE, P.J., DIXON, L. and URQUHART, A (2001) Sequence polymorphism in the human melanocortin 1 receptor gene as an indicator of the red hair phenotype. *Forensic science international,* 122(2): 124-129.

GRUBWEISER P, MUHLMANN R, PARSON W (2006) A new "miniSTR-multiplex" displaying reduced amplicon lengths for the analysis of degraded DNA. *International Journal of Legal Medicine*, 120(9): 115-120.

GUNTER, C (2005) She moves in mysterious ways. *Nature,* 434(7031): 279-280.

HALL, T. A (2011) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. http://www.mbio.ncsu.edu/BioEdit/bioedit.html.

HARRIS, K.A., THACKER, C.R., BALLARD, D. and COURT, D.S (2006) *The effect of cleaning agents on the DNA analysis of blood stains deposited on different substrates. International Congress Series, 1288 (2): 589-591.*

HARTZELL, B., GRAHAM, K. and MCCORD, B (2003) Response of short tandem repeat systems to temperature and sizing methods. *Forensic Science International,* 133(3): 228-234.

HELLMANN, A., ET AL (2001) STR typing of human telogen hairs — A new approach. *International Journal of Legal Medicine, 114*(4): 269 – 273.

HILL C.R., and BUTLER, J.M (2010) An overview of ABI 3500 Genetic Analyzer. *American Society of Forensic Science,* 133(3): 556-589

HILL C., KLINE M.C., MULERO J.J., LAGACE R.E., CHANGE C.W (2007) Concordance Study Between the AmpFlSTR MiniFiler PCR Amplification Kit and Conventional STR Typing Kits. *Journal of Forensic Sciences*, 52(2): 870-873.

HILL, C.R., KLINE, M.C., COBLE, M.D. and BUTLER, J.M (2008) Characterization of 26 miniSTR loci for improved analysis of degraded DNA samples. *Journal of Forensic Sciences,* 53(1): 73-80.

HOFREITER, M., SERRE, D., POINAR, H. N., KUCH, M. & PAABO, S (2001) Ancient DNA. *Nature, 44(*2): 353-359.

HOLLAND, M. M., CAVE, C. A., HOLLAND, C. A. & BILLE, T. W (2003) Development of a quality, high throughput DNA analysis procedure for skeletal samples to assist with the identification of victims from the World Trade Center attacks. *Croatian Medical Journal,* 44(3): 264-72.

HOLLAND, M and PARSON, W (1999) The future of evolutionary developmental biology. *Nature,* 402(6): 41-44.

HOLLARD, C., MENDISCO, F., KEYSER, C., CRUBÉZY, E. and LUDES, B (2011) First application of the Investigator DIPplex indels typing kit for the analysis of ancient DNA samples. *Forensic Science International: Genetics,* 3(1): 393-394.

HOOT, S.B. and DOUGLAS, A.W (1998) Phylogeny of the Proteaceae based on atpB and atpB-rbcL intergenic spacer region sequences. *Australian Systematic Botany,* 11(3): 301-320.

HUANG, C.R.L., SCHNEIDER, A.M., LU, Y., NIRANJAN, T., SHEN, P., ROBINSON, M.A., STERANKA, J.P., VALLE, D., CIVIN, C.I., WANG, T., WHEELAN, S.J., JI, H., BOEKE, J.D. and BURNS, K.H (2010) Mobile interspersed repeats are major structural variants in the human genome. *Cell,* 141(7): 1171-1182.

HUDLOW, W.R., KRIEGER, R., MEUSEL, M., SEHHAT, J.C., TIMKEN, M.D. & BUONCRISTIANI, M.R (2011) The NucleoSpin (R) DNA Clean-up XS kit for the concentration and purification of genomic DNA extracts: An alternative to microdialysis filtration. *Forensic Science International-Genetics,* 5(3): 226-230.

IOSHIKHES, I.P., ALBERT, I., ZANTON, S.J. and PUGH, B.F (2006) Nucleosome positions predicted through comparative genomics. *Nature Genetics,* 38(10): 1210-1215.

ISENBERG, A.R (2004) Forensic mitochondrial DNA analysis. *Forensic Science International, 2(2):* 297 – 327.

ISKOW, R.C., MCCABE, M.T., MILLS, R.E., TORENE, S., PITTARD, W.S., NEUWALD, A.F., VAN MEIR, E.G., VERTINO, P.M. and DEVINE, S.E (2010) Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell,* 141(7): 1253-1261.

ITANI M, Y. Y., DOI Y, MIYAISHI, S (2011) Quantitative analysis of DNA degradation in the dead body. *Acta Medica Okayama,* 65(2): 299-306.

JARMAN, A.P., NICHOLLS, R.D., WEATHERALL, D.J., CLEGG, J.B. and HIGGS, D.R (1986) Molecular characterisation of a hypervariable region downstream of the human alpha-globin gene cluster. *The EMBO journal,* 5(8): 1857-1863.

JEFFREYS, A. J., WILSON, V. & THEIN, S. L (1985) Hypervariable 'minisatellite' regions in human DNA. *Nature,* 314(3): 67-73.

JOBLING, M, HOLLOX, E., HURLES,M., KIVISILD, T. and TYLER-SMITH, C (2013) *Human Evolutionary Genetics*, 2nd Edition, Garland Science.

JOBLING, M. A., GILL, P (2004) Encoded evidence: DNA in forensic analysis. *Nature Genetics,* 5(10): 739-751.

JOBLING, M.A (2001) Human pigmentation: Not all black and white. *Trends in Genetics,* 17(11): 625-630.

KARLSSON, A.O., HOLMLUND, G., EGELAND, T. and MOSTAD, P (2007) DNA-testing for immigration cases: The risk of erroneous conclusions. *Forensic Science International,* 172(2): 144-149.

KATCHER, H. & SCHWARTZ, I (1994). A Distinctive Property of the DNA-Polymerase - Enzymatic Amplification in the Presence of Phenol. *BioTechniques,* 16(1): 84-85.

KAYSER, M. & DE KNIFF, P (2011) Improving human forensics through advances in genetics, genomics and molecular biology. *Nature Genetics*, 12(1): 179-192.

KAYSER, M (2015) Forensic DNA Phenotyping: Predicting human appearance from crime scene material for investigative purposes. *Forensic Science International: Genetics,* 18(2): 33-48.

KAYSER, M., VERMEULEN, M., KNOBLAUCH, H., SCHUSTER, H., KRAWCZAK, M. and ROEWER, L (2007) Relating two deep-rooted pedigrees from Central Germany by high-resolution Y-STR haplotyping. *Forensic Science International: Genetics,* 1(2): 125-128.

KERSBERGEN, P. VAN DUIJN, A.D. KLOOSTERMAN, J.T. DEN DUNNEN, T. KAYSER, M. KNIJFF, P. D (2009) Developing a set of ancestry-sensitive DNA markers reflecting continental origins of humans. *BMC Genetics*, 10(2): 69-71.

KIDD, J.M., COOPER, G.M., DONAHUE, W.F., HAYDEN, H.S., SAMPAS, N., GRAVES, T., HANSEN, N., TEAGUE, B., ALKAN, C., ANTONACCI, F., HAUGEN, E., ZERR, T., YAMADA, N.A., TSANG, P., NEWMAN, T.L., TÜZÜN, E., CHENG, Z., EBLING, H.M., TUSNEEM, N., DAVID, R., GILLETT, W., PHELPS, K.A., WEAVER, M., SARANGA, D., BRAND, A., TAO, W., GUSTAFSON, E., MCKERNAN, K., CHEN, L., MALIG, M., SMITH, J.D., KORN, J.M., MCCARROLL, S.A., ALTSHULER, D.A., PEIFFER, D.A., DORSCHNER, M., STAMATOYANNOPOULOS, J., SCHWARTZ, D., NICKERSON, D.A., MULLIKIN, J.C., WILSON, R.K., BRUHN, L., OLSON, M.V., KAUL, R., SMITH, D.R. and EICHLER, E.E (2008) Mapping and sequencing of structural variation from eight human genomes. *Nature,* 453(7191): 56-64.

KIM, S., LABBE, R. & RYU, S (2000) Inhibitory effects of collagen on the PCR for detection of Clostridium perfringens. *Applied and Environmental Microbiology,* 66(3): 1213-1215.

KING, T.E. E.J. PARKIN, G. SWINFIELD, F. CRUCIANI, R. SCOZZARI, A. ROSA, S.K. LIM, Y. XUE, C. TYLER-SMITH, M.A. JOBLING,A (2007) The deepest-rooting clade of the Y phylogeny within an English genealogy. *European Journal of Human Genetics*, 15(3): 288–293.

KIS, Z., ZALÁN, A., VÖLGYI, A., KOZMA, Z., DOMJÁN, L. and PAMJAV, H (2012) Genome deletion and insertion polymorphisms (DIPs) in the Hungarian population. *Forensic Science International: Genetics,* 6(5): 125-126.

KLINE, M., DUEWER, D., REDMAN, J. & BUTLER, J (2005) Results from the NIST 2004 DNA quantitation study. *Journal of Forensic Sciences,* 50(3): 571-578.

KRENKE B., TEREBA A., ANDERSON S., BUEL E., CULHANE S., FINIS C (2002) Validation of a 16-locus fluorescent multiplex system. *Journal of Forensic Sciences,* 47(2): 773-785.

LAMASON, R.L., MOHIDEEN, M.-.P.K., MEST, J.R., WONG, A.C., NORTON, H.L., AROS, M.C., JURYNEC, M.J., MAO, X., HUMPHREVILLE, V.R., HUMBERT, J.E., SINHA, S., MOORE, J.L., JAGADEESWARAN, P., ZHAO, W., NING, G., MAKALOWSKA, I., MCKEIGUE, P.M., O'DONNELL, D., KITTLES, R., PARRA, E.J., MANGINI, N.J., GRUNWALD, D.J., SHRIVER, M.D., CANFIELD, V.A. and CHENG, K.C (2005) Genetics: SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science,* 310(5755): 1782-1786.

LANDSTEINER, K (1900) Zur Kenntnis der antifermentativen, lytisichen und agglutinierenden Wirkungen des Blutserums und der Lymphe. *Zentralbl. Bakteriol*, 27(2): 357-362.

LANTZ, P., MATSSON, M., WADSTRÖM, T. and RÅDSTRÖM, P (1997) Removal of PCR inhibitors from human faecal samples through the use of an aqueous two-phase system for sample preparation prior to PCR. *Journal of Microbiological Methods,* 28(3): 159-167.

LARKIN, A. & HARBISON, S (1999) An improved method for STR analysis of bloodstained denim. *International Journal of Legal Medicine,* 112(6): 388-390.

LARKIN, B., IASCHI, S., DADOUR, I. & TAY, G (2010) Using accumulated degree-days to estimate post-mortem interval from the DNA yield of porcine skeletal muscle. *Forensic Science, Medicine, and Pathology,* 6(1): 83-92.

LARUE, B.L., GE, J., KING, J.L. and BUDOWLE, B (2012) A validation study of the Qiagen Investigator DIPplex® kit; An INDEL-based assay for human identification. *International Journal of Legal Medicine,* 126(4): 533-540.

LARUE, B.L., LAGACÉ, R., CHANG, C., HOLT, A., HENNESSY, L., GE, J., KING, J.L., CHAKRABORTY, R. and BUDOWLE, B (2014) Characterization of 114 insertion/deletion (INDEL) polymorphisms, and selection for a global INDEL panel for human identification. *International Journal of Legal medicine, 33*(2): 55-58.

LEE, H.Y., PARK, M.J., KIM, N.Y., SIM, J.E., YANG, W.I. & SHIN, K (2010) Simple and highly effective DNA extraction methods from old skeletal remains using silica columns. *Forensic Science International-Genetics,* 4(5): 275-280.

LEVINSON, G. and GUTMAN, G.A (1987) Slipped-strand mispairing; A major mechanism for DNA sequence evolution. *Molecular Biology and Evolution,* 4(3): 203-221.

LEHMANN, A.R (2003) Replication of damaged DNA. *Cell,* 2(4): 300-302.

LEVINSON, G., & GUTMAN, G. A (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology Evolution*, 4(3): 203-221.

LEWONTIN, R. C (1972) The apportionment of human diversity. *Molecular Biology Evolution,* 6(1): 381-398.

LI, J.Z., ABSHER, D.M., TANG, H., SOUTHWICK, A.M., CASTO, A.M., RAMACHANDRAN, S., CANN, H.M., BARSH, G.S., FELDMAN, M., CAVALLI-SFORZA, L.L. and MYERS, R.M (2008) Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation. *Science*, 319(5866): 1100-1104.

LI, C., ZHAO, S., ZHANG, S., LI, L., LIU, Y., CHEN, J. and XUE, J (2011) Genetic polymorphism of 29 highly informative InDel markers for forensic use in the Chinese Han population. *Forensic Science International: Genetics,* 5(1): 27-30.

LIANG, W., ZAUMSEGEL, D., ROTHSCHILD, M.A., LV, M., ZHANG, L., LI, J., LIU, F., XIANG, J. and SCHNEIDER, P.M (2013) Genetic data for 30 insertion/deletion polymorphisms in six Chinese populations with Qiagen Investigator DIPplex Kit. *Forensic Science International: Genetics Supplement Series,* 4(1): 268-269.

LINDAHL, T (1993) Instability and decay of the primary structure of DNA. *Nature,* 362(8): 709-715.

LONDIN, E.R. M.A. KELLER, C. MAISTA, G. SMITH, L.A. MAMOUNAS, R. ZHANG, S.J. MADORE, K. GWINN, R.A. CORRIVEAU (2010) A cost-effective panel of ancestry informative markers for determining continental origins. *PLoS One*, 5(1): 134-143.

LOPES, T.R., SANTOS, S., RIBEIRO-DOS-SANTOS, Â., RESQUE, R.L., PINTO, G.R. and YOSHIOKA, F.K.N (2014) Population data of the 46 insertion–deletion (INDEL) loci in population in Piauí State, Northeastern Brazil. *Forensic Science International: Genetics,* 9(2): 13-15.

LUCE, C., MONTPETIT, S., GANGITANO, D. and O'DONNELL, P (2009) Validation of the AMPFlSTR® MiniFiler PCR amplification kit for use in forensic casework. *Journal of Forensic Sciences,* 54(5): 1046-1054.

LUNDSBERG, B., JOHANSEN, P., BØRSTING, C. and MORLING, N (2013) Development and optimisation of five multiplex assays with 115 of the AIM SNPs from the EUROFORGEN AIMs set on the Sequenom® MassARRAY® system. *Forensic Science International: Genetics Supplement Series,* 4(1): 182-183.

LUPSKI, J.R (2013) Genome mosaicism - One human, multiple genomes. *Science,* 341(6144): 358-359.

MAHONY, J., CHONG, S., JANG, D., LUINSTRA, K., FAUGHT, M., DALBY, D., SELLORS, J. and CHERNESKY, M (1998) Urine specimens from pregnant and non-pregnant women inhibitory to amplification of Chlamydia trachomatis nucleic acid by PCR, ligase chain reaction, and transcription-mediated amplification: Identification of urinary substances associated with inhibition and removal of inhibitory activity. *Journal of Clinical Microbiology,* 36(11): 3122-3126.

MARJANOVIC, D., DURMIC-PASIC, A., KOVACEVIC, L., AVDIC, J., DZEHVEROVIC, M., HAVERIC, S., RAMIC, J., KALAMUJIC, B., BILELA, L.L., SKARO, V., PROJIC, P., BAJROVIC, K., DROBNIC, K., DAVOREN, J. & PRIMORAC, D (2009) Identification of Skeletal Remains of Communist Armed Forces Victims During and After World War II: Combined Y-chromosome Short Tandem Repeat (STR) and MiniSTR Approach. *Croatian Medical Journal,* 50(3): 296-304.

MANTA, F., CAIAFA, A., PEREIRA, R., SILVA, D., AMORIM, A., CARVALHO, E.F. and GUSMÃO, L (2012) Indel markers: Genetic diversity of 38 polymorphisms in Brazilian populations and application in a paternity investigation with post mortem material. *Forensic Science International: Genetics,* 6(5): 658-661.

MANTA, F.S.N., PEREIRA, R., CAIAFA, A., SILVA, D.A., GUSMÃO, L. and CARVALHO, E.F (2013) Analysis of genetic ancestry in the admixed Brazilian population from Rio de Janeiro using 46 autosomal ancestry-informative indel markers. *Annals of Human Biology,* 40(1): 94-98.

MARTÍN, P., GARCÍA, O., HEINRICHS, B., YURREBASO, I., AGUIRRE, A. and ALONSO, A (2013) Population genetic data of 30 autosomal indels in Central Spain and the Basque Country populations. *Forensic Science International: Genetics,* 7(2): 27-30.

MARTÍNEZ-CORTÉS, G., GUSMÃO, L., PEREIRA, R., SALCIDO, V.H., FAVELA-MENDOZA, A.F., MUÑOZ-VALLE, J.F., INCLÁN-SÁNCHEZ, A., LÓPEZ-HERNÁNDEZ, L.B. and RANGEL-VILLALOBOS, H (2015) Genetic structure and forensic parameters of 38 Indels for human identification purposes in eight Mexican populations. *Forensic Science International: Genetics,* 17(4): 149-152.

MCDONALD, J.P., HALL, A., GASPARUTTO, D., CADET, J., BALLANTYNE, J. and WOODGATE, R (2006) Novel thermostable Y-family polymerases: Applications for the PCR amplification of damaged or ancient DNAs. *Nucleic acids research,* 34(4): 1102-1111.

MESSER, P.W (2007) DNA Insertions and Deletions in the Human Genome. *Molecular Genetics,* 33(2): 34-36.

MILLS, R.E., BENNETT, E.A., ISKOW, R.C. and DEVINE, S.E (2007) Which transposable elements are active in the human genome? *Trends in Genetics,* 23(4): 183-191.

MILLS, R.E., LUTTIG, C.T., LARKINS, C.E., BEAUCHAMP, A., TSUI, C., PITTARD, W.S. and DEVINE, S.E (2006) An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Research,* 16(9): 1182-1190.

MILLS, R.E., PITTARD, W.S., MULLANEY, J.M., FAROOQ, U., CREASY, T.H., MAHURKAR, A.A., KEMEZA, D.M., STRASSLER, D.S., PONTING, C.P., WEBBER, C. and DEVINE, S.E (2011) Natural genetic variation caused by small insertions and deletions in the human genome. *Genome Research,* 21(6): 830-839.

MONICA, A., PRUGNOLLE, F., & BALLOUX, F (2005) Geography is a better determinant of human genetic differentiation than ethnicity. *Human Genetics,* 118(3): 366-371.

MONTGOMERY, S.B., GOODE, D.L., KVIKSTAD, E., ALBERS, C.A., ZHANG, Z.D., MU, X.J., ANANDA, G., HOWIE, B., KARCZEWSKI, K.J., SMITH, K.S., ANAYA, V., RICHARDSON, R., DAVIS, J., MACARTHUR, D.G., SIDOW, A., DURET, L., GERSTEIN, M., MAKOVA, K.D., MARCHINI, J., MCVEAN, G. and LUNTER, G (2013) The origin, evolution, and functional impact of short insertion-deletion variants identified in 179 human genomes. *Genome Research,* 23(5): 749-761.

MONTEIRO, L., BONNEMAISON, D., VEKRIS, A., PETRY, K.G., BONNET, J., VIDAL, R., CABRITA, J. and MÉGRAUD, F (1997) Complex polysaccharides as PCR inhibitors in feces: Helicobacter pylori model. *Journal of Clinical Microbiology,* 35(4): 995-998.

MORETTI TR, BAUMSTARK AL, DEFENBAUGH DA, KEYS KM, SMERICK JB, BUDOWLE B (2001) Validation of short tandem repeats (STRs) for forensic usage: performing testing of fluorescent multiplex STR systems and analysis of authentic and simulated forensic samples. *Journal of Forensic Sciences*, 46(3): 647-660.

MOREIRA D (1998) Efficient removal of PCR inhibitors using agarose-embedded DNA preparations. *Nucleic Acids Research*, 26(3): 3309-3319.

MULERO, J. J., CHANG, C. W., LAGACÉ, R. E., WANG, D. Y., BAS, J. L., MCMAHON, T. P. & HENNESSY, L. K (2008) Development and Validation of the AmpFℓSTR® MiniFilerTM PCR Amplification Kit: A MiniSTR Multiplex for the Analysis of Degraded and/or PCR Inhibited DNA. *Journal of Forensic Sciences,* 53(2): 838-852.

MÜLLER, K., KLEIN, R., MILTNER, E. and WIEGAND, P (2007) Improved STR typing of telogen hair root and hair shaft DNA. *Electrophoresis,* 28(16): 2835-2842.

MÜLLER, K., SOMMERER, T., MILTNER, E., SCHNEIDER, H. and WIEGAND, P (2010) Casework testing of the multiplex kits AmpFlSTR® SEfiler Plus™ PCR amplification kit (AB), PowerPlex® S5 System (Promega) and AmpFlSTR® MiniFiler™ PCR amplification kit (AB). *Forensic Science International: Genetics,* 4(3): 200-205.

MULLIKLN, J.C., HUNT, S.E., COLE, C.G., MORTIMORE, B.J., RICE, C.M., BURTON, J., MATTHEWS, L.H., PAVITT, R., PLUMB, R.W., SIMS, S.K., ALNSCOUGH, R.M.R., ATTWOOD, J., BALLEY, J.M., BARLOW, K., BRUSKIEWICH, R.M.M., BUTCHER, P.N., CARTER, N.P., CHEN, Y., CLEE, C.M., COGGILL, P.C., DAVIES, J., DAVIES, R.M., DAWSON, E., FRANCIS, M.D., JOY, A.A., LAMBLE, R.G., LANGFORD, C.F., MACARTHY, J., MALL, V., MORELAND, A., OVERTON-LARTY, E.K., ROSS, M.T., SMITH, L.C., STEWARD, C.A., SULSTON, J.E., TINSLEY, E.J., TURNEY, K.J., WILLEY, D.L., WILSON, G.D., MCMURRAY, A.A., DUNHAM, I., ROGERS, J. and BENTLEY, D.R (2000) An SNP map of human chromosome 22. *Nature,* 407(6): 516-520.

MULLIS, K., FALOONA, F., SCHARF, S., SAIKI, R., HORN, G., & ERLICH, H (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb. Symp. *Quantitative Biology*, 51(1): 263-273.

MYLES, S., STONEKING, M. and TIMPSON, N (2009) An assessment of the portability of ancestry informative markers between human populations. *BMC Medical Genomics, 44(*2): 547-551.

NACHMAN, M.W. and CROWELL, S.L (2000) Estimate of the mutation rate per nucleotide in humans. *Genetics,* 156(1): 297-304.

NAKAMURA, Y., LEPPERT, M., O'CONNELL, P., WOLFF, R., HOLM, T., CULVER, M., MARTIN, C., FUJIMOTO, E., HOFF, M., KUMLIN, E., & WHITE, R (1987) Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science*, 235(2): 1616-1622.

NELSON, J (2009) Repair of Damaged DNA for Forensic Analysis. *Science*, 34(3): 454-450.

NELSON, K., and MELTON, T (2007) Forensic mitochondrial DNA analysis of 116 casework skeletal samples. *Journal of Forensic Sciences,* 52(3): 557-561.

NEUVONEN, A.M., PALO, J.U., HEDMAN, M. and SAJANTILA, A (2011) Discrimination power of Investigator DIPplex loci in Finnish and Somali populations. *Forensic Science International: Genetics,* 6(4): 99-102.

NIEDERSTÄTTER, H., COBLE, M.D., GRUBWIESER, P., PARSONS, T.J. and PARSON, W (2006) Characterization of mtDNA SNP typing and mixture ratio assessment with simultaneous real-time PCR quantification of both allelic states. *International Journal of Legal Medicine,* 120(1): 18-23.

NOREN, L., HEDELL, R., ANSELL, R. & HEDMAN, J (2013) Purification of crime scene DNA extracts using centrifugal filter devices. *Investigative Genetics,* 4(1): 8-10.

NUNOTANI, M., SHIOZAKI, T., SATO, N., KAMEI, S., TAKATSU, K., HAYASHI, T., OTA, M. and ASAMURA, H (2015) Analysis of 30 insertion–deletion polymorphisms in the Japanese population using the Investigator DIPplex® kit. *International Journal of Legal Medicine,* 17(6): 467-470.

O'CONNOR, K.L., HILL, C.R., VALLONE, P.M. and BUTLER, J.M (2011) Linkage disequilibrium analysis of D12S391 and vWA in U.S. population and paternity samples. *Forensic Science International: Genetics,* 5(5): 538-540.

O'CONNOR, K.L. and TILLMAR, A.O (2012) Effect of linkage between vWA and D12S391 in kinship analysis. *Forensic Science International: Genetics,* 6(6): 840-844.

OKA, A., TAKADA, T., FUJISAWA, H. and SHIROISHI, T (2014) Evolutionarily Diverged Regulation of X-chromosomal Genes as a Primal Event in Mouse Reproductive Isolation. *PLoS Genetics,* 10(4): 345-356.

OLDONI, F., CASTELLA, V. and HALL, D (2015) A novel set of DIP-STR markers for improved analysis of challenging DNA mixtures. *Forensic Science International: Genetics,* 19(2): 156-164.

OPEL KL, CHUNG D, DRABEK J, TATAREK N, JANTZ L, AND MCCORD B (2006) The Application of Miniplex Primer Sets in the Analysis of Degraded DNA from Human Skeletal Remains. *Journal of Forensic Sciences*, 51(2): 351-356.

OPEL K.L., CHUNG D.T., DRABEK J., BUTLER J.M., MCCORD B.R (2008) Developmental Validation of Reduced-Size STR Miniplex Primer Sets. *Journal of Forensic Science,* 52(6): 1263-1271.

PÄÄBO, S., POINAR, H., SERRE, D., JAENICKE-DESPRÉS, V., HEBLER, J., ROHLAND, N., KUCH, M., KRAUSE, J., VIGILANT, L. & HOFREITER, M (2004) Genetic analysis from ancient DNA. *Annual Review of Genetics,* 38(3): 645-679.

PALO, J.U., PIRTTIMAA, M., BENGS, A., JOHNSSON, V., ULMANEN, I., LUKKA, M., UDD, B. and SAJANTILA, A (2007) The effect of number of loci on geographical structuring and forensic applicability of Y-STR data in Finland. *International Journal of Legal Medicine,* 122(6): 449-456.

PARSONS, T. J (2007) Application of novel ‘ mini-amplicon ’ STR multiplexes to high volume casework on degraded skeletal remains . *Forensic Science International: Genetics, 45(1):* 175 – 179.

PARSON, W (2007) EMPOP—a forensic mtDNA database. Forensic Science International: Genetics, 34(1): 88–92.

PASSI, N., GARG, R.K., YADAV, M., SINGH, R.S. and KHAROSHAH, M.A (2012) Effect of luminol and bleaching agent on the serological and DNA analysis from bloodstain. *Egyptian Journal of Forensic Sciences,* 2(2): 54-61.

PATZELT, D (2004) History of forensic serology and molecular genetics in the sphere of activity of the German Society for Forensic Medicine. *Forensic science international,* 144(2): 185-191.

PEPINSKI, W., ABREU-GLOWACKA, M., KORALEWSKA-KORDEL, M., MICHALAK, E., KORDEL, K., NIEMCUNOWICZ-JANICA, A., SZEREMETA, M. and KONARZEWSKA, M (2013) Population genetics of 30 INDELs in populations of Poland and Taiwan. *Molecular Biology Reports,* 40(7): 4333-4338.

PEREGRINE, P.N., EMBER, C.R. and EMBER, M (2003) Cross-cultural evaluation of predicted associations between race and behavior. *Evolution and Human Behavior,* 24(5): 357-364.

PEREIRA, R., PHILLIPS, C., ALVES, C., AMORIM, A., CARRACEDO, Á and GUSMAO, L (2009) A new multiplex for human identification using insertion/deletion polymorphisms. *Electrophoresis,* 30(21): 3682-3690.

PEREIRA, R., PHILLIPS, C., PINTO, N., SANTOS, C., DOS SANTOS, S.E.B., AMORIM, A., CARRACEDO, A. and GUSMÃO, L (2012) Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. *PLoS ONE,* 7(1): 458-468.

PFAFF, C.L., BARNHOLTZ-SLOAN, J., WAGNER, J.K. and LONG, J.C (2004) Information on Ancestry from Genetic Markers. *Genetic Epidemiology,* 26(4): 305-315.

PHILLIPS, C., SALAS, A., SANCHEZ, J. J., FONDEVILA, M., GOMEZ-TATO, A., ALVAREZ-DIOS, J., CALAZA, M., DE CAL, M. C., BALLARD, D., LAREU, M. V., CARRACEDO, A. & SNPFORID CONSORTIUM (2007) Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. *Forensic Science International: Genetics,* 33(1): 273-280.

PHILLIPS, C. FERNANDEZ-FORMOSO, M. GARCIAÑAS, L. PORRAS, T. TVEDEBRINK, J. AMIGO, M. FONDEVILA, A. GOMEZ-TATO, J. ALVAREZ-DIOS, A. FREIRE-ARADAS, D (2011) Analysis of global variability in 15 established and 5 new European Standard Set (ESS) STRs using the CEPH human genome diversity panel. *Forensic Science International: Genetics,* 56(5): 155–169.

PINTO, N., MAGALHÃES, M., CONDE-SOUSA, E., GOMES, C., PEREIRA, R., ALVES, C., GUSMÃO, L. and AMORIM, A (2013) Assessing paternities with inconclusive STR results: The suitability of bi-allelic markers. *Forensic Science International: Genetics,* 7(1): 16-21.

PRINCE, A.M. and ANDRUS, L (1992) PCR: How to kill unwanted DNA. *BioTechniques,* 12(3): 233-245.

PRITCHARD, J. K., STEPHENS, M., & DONNELLY, P (2000) Inference of population structure using multilocus genotype data. *Genetics,* 155 (2): 945-959.

PRITCHARD, J. K., WEN, X., & FALUSH, D (2010) Documentation for STRUCTURE software: version 2.3'. [online] Available at: http://pritch.bsd.uchicago.edu/structure.html

RÅDSTRÖM, P., KNUTSSON, R., WOLFFS, P., LÖVENKLEV, M. and LÖFSTRÖM, C (2004) Pre-PCR processing: Strategies to generate PCR-compatible samples. *Applied Biochemistry and Biotechnology,* 26(2): 133-146.

RADWAN, A., YOUNIS, A., LUYKX, P. and KHURI, S (2008) Prediction and analysis of nucleosome exclusion regions in the human genome. *BMC Genomics,* 9(3): 457-462.

RAMACHANDRAN, S., DESHPANDE, O., ROSEMAN, C. C., ROSENBERG, N. A., FELDMAN, M. W., and CAVALLI-SFORZA, L. L (2005) Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *National Academy of Science. USA* 102(4): 15942–15947.

RAYMOND, M. and ROUSSET, F (1995) An exact test for population differentiation. *Evolution*, 49(6): 1280-1283

ROBINS J, M. S. E., FUREY L (2001) Hit or miss? Factors affecting DNA preservation in Pacific and archaeological material. *Australian Archaemetry,* 57(4): 478-489.

ROCKENBAUER, E., HANSEN, S., MIKKELSEN, M., BØRSTING, C. and MORLING, N (2014) Characterization of mutations and sequence variants in the D21S11 locus by next generation sequencing. *Forensic Science International: Genetics,* 8(1): 68-72.

RODIG, H., ROEWER, L., GROSS, A., RICHTER, T., DE KNIJFF, P., KAYSER, M. and BRABETZ, W (2008) Evaluation of haplotype discrimination capacity of 35 Y-chromosomal short tandem repeat loci. *Forensic Science International,* 174(2): 182-188.

ROHLFS, R.V. FULLERTON, S.M. WEIR, B.S (2012) Familial identification: population structure and relationship distinguishability. *PLoS Genetics,* 8(3): 458-462.

ROMANINI, C., CATELLI, M.L., BOROSKY, A., PEREIRA, R., ROMERO, M., SALADO PUERTO, M., PHILLIPS, C., FONDEVILA, M., FREIRE, A., SANTOS, C., CARRACEDO, A., LAREU, M.V., GUSMAO, L. and VULLO, C.M (2012). Typing short amplicon binary polymorphisms: Supplementary SNP and Indel genetic information in the analysis of highly degraded skeletal remains. *Forensic Science International: Genetics,* 6(4): 469-476.

ROMANINI, C., ROMERO, M., SALADO PUERTO, M., CATELLI, L., PHILLIPS, C., PEREIRA, R., GUSMÃO, L. and VULLO, C (2014) Ancestry informative markers: Inference of ancestry in aged bone samples using an autosomal AIM-Indel multiplex. *Forensic Science International: Genetics,* 16(3): 58-63.

ROSENBERG, N. A., MAHAJAN, S., RAMACHANDRAN, S., ZHAO, C., PRITCHARD, J. K (2005) Clines, clusters, and the effect of study design on the inference of human population structure. *PLoS Genetics,* 1(6): 70-75.

ROSENBERG, N. A., PRITCHARD, J. K., WEBER J. L., CANN, H. M., KIDD, K. K., ZHIVOTOVSKY L. A., FELDMAN, M. W (2002) Genetic structure of human populations. *Science,* 298(4): 2381-2385.

ROSS, M.T., GRAFHAM, D.V., COFFEY, A.J., SCHERER, S., MCLAY, K (2005) The DNA sequence of the human X chromosome. *Nature,* 434(7031): 325-337.

ROSSEN, L., NORSKOV, P., HOLMSTROM, K. & RASMUSSEN, O (1992) Inhibition of PCR by Components of Food Samples, Microbial Diagnostic Assays and DNA-Extraction Solutions. *International Journal of Food Microbiology,* 17(1): 37-45.

ROWOLD, D.J., LUIS, J.R., TERREROS, M.C. and HERRERA, R.J (2007) Mitochondrial DNA geneflow indicates preferred usage of the Levant Corridor over the Horn of Africa passageway. *Journal of Human Genetics,* 52(5): 436-447.

SAIKI, R.K., GELFAND, D.H., STOFFEL, S., SCHARF, S.J., HIGUCHI, R., HORN, G.T., MULLIS, K.B. and ERLICH, H.A (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science,* 239(4839): 487-491.

SAIZ, M., ANDRÉ, F., PISANO, N., SANDBERG, N., BERTONI, B. and PAGANO, S (2012) Allelic frequencies and statistical data from 30 INDEL loci in Uruguayan population. *Forensic Science International: Genetics, 45*(1): 234-241.

SAIZ, M., ALVAREZ-CUBERO, M.J., MARTINEZ-GONZALEZ, L.J., ALVAREZ, J.C. and LORENTE, J.A (2014) Population genetic data of 38 insertion–deletion markers in South East Spanish population. *Forensic Science International: Genetics,* 13(1): 236-238.

SAJANTILA A, PUOMILAHTI S, JOHNSSON V, EHNHOLM C (1992) Amplification of reproducible allele markers for amplified fragment length polymorphism analysis. *Biotechniques,* 12(2): 20-22.

SALAS, A., BANDELT, H.-J., MACAULAY, V. & RICHARDS, M. B (2007) Phylogeographic investigations: The role of trees in forensic genetics. *Forensic Science International,* 168(1): 01-13.

SANCHEZ, A.M., VOLK, D.E., GORENSTEIN, D.G. and LLOYD, R.S (2003) Initiation of repair of A/G mismatches is modulated by sequence context. *DNA Repair,* 2(8): 863-878.

SANCHEZ, J. J. & ENDICOTT, P (2006) Developing multiplexed SNP assays with special reference to degraded templates. *Nature, 34(*1): 1370-1378.

SANTA RITA, T.H., CHIANCA, C.F., NERY, L.F.A. and BARRA, G.B (2015) Allele frequencies of 38 insertion/deletion polymorphisms in a population sample from the Federal District (Brazil)—A territory that arose from nothing. *Forensic Science International: Genetics,* 19(2): 79-80.

SANTOS, C., FONDEVILA, M., BALLARD, D., BANEMANN, R., BENTO, A.M., BØRSTING, C., BRANICKI, W., BRISIGHELLI, F., BURRINGTON, M., CAPAL, T., CHAITANYA, L., DANIEL, R., DECROYER, V., ENGLAND, R., GETTINGS, K.B., GROSS, T.E., HAAS, C., HARTEVELD, J., HOFF-OLSEN, P., HOFFMANN, A., KAYSER, M., KOHLER, P., LINACRE, A., MAYR-EDUARDOFF, M., MCGOVERN, C., MORLING, N., O'DONNELL, G., PARSON, W., PASCALI, V.L., PORTO, M.J., ROSETH, A., SCHNEIDER, P.M., SIJEN, T., STENZL, V., COURT, D.S., TEMPLETON, J.E., TURANSKA, M., VALLONE, P.M., OORSCHOT, R.A.H.V., ZATKALIKOVA, L., CARRACEDO, Á. and PHILLIPS, C (2015) Forensic ancestry analysis with two capillary electrophoresis ancestry informative marker (AIM) panels: Results of a collaborative EDNAP exercise. *Forensic Science International: Genetics,* 19(1): 56-67.

SANTOS, N.P.C., RIBEIRO-RODRIGUES, E.M., RIBEIRO-DOS-SANTOS, ÂK.C., PEREIRA, R., GUSMÃO, L., AMORIM, A., GUERREIRO, J.F., ZAGO, M.A., MATTE, C., HUTZ, M.H. and DOS SANTOS, S.E.B (2010) Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INSEL) ancestry-informative marker (AIM) panel. *Human Mutation,* 31(2): 184-190.

SCHOLZ M, GIDDINGS I, PUSCH CM (1998) A polymerase chain reaction inhibitor of ancient hard and soft tissue DNA extracts is determined as human collagen type I. *Anal Biochemistry,* 259(283): 286-287.

SCHULZ MM, REICHERT W (2002) *Archived or directly swabbed latent fingerprints as a DNA source for STR typing. Forensic Science International, 1*27(2): 128-130.

SCHMERER, W. M., HUMMEL, S., & HERRMANN, B (1999) Optimized DNA extraction to improve reproducibility of short tandem repeat genotyping with highly degraded DNA as target. *Electrophoresis,* 20(1): 1712-1716.

SCHRADER, C., SCHIELKE, A., ELLERBROEK, L. and JOHNE, R (2012) PCR inhibitors - occurrence, properties and removal. *Journal of Applied Microbiology,* 113(5): 1014-1026.

SCHUMM, J.W., WINGROVE, R.S. and DOUGLAS, E.K (2004) Robust STR multiplexes for challenging casework samples. *International Congress Series,* 1261(1): 547-549.

SEONG, K.M., PARK, J.H., HYUN, Y.S., KANG, P.W., CHOI, D.H., HAN, M.S., PARK, K.W. and CHUNG, K.W (2014) Population genetics of insertion–deletion polymorphisms in South Koreans using Investigator DIPplex kit. *Forensic Science International: Genetics,* 8(1): 80-83.

SERRE D, HOFREITER M AND PÄÄBO S (2004) Mutations Induced by Ancient DNA Extracts. *Molecular Biology Evolution,* 21(8): 1463–1467.

SHRIVER, M.D., KENNEDY, G.C., PARRA, E.J., LAWSON, H.A., SONPAR, V., HUANG, J., AKEY, J.M. and JONES, K.W (2004) The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. *Human Genomics,* 1(4): 274-286.

SHRIVER, M.D. and KITTLES, R.A (2004) Genetic ancestry and the search for personalized genetic histories. *Nature Reviews Genetics,* 5(8): 611-618.

SHRIVER, M.D. SMITH, L. JIN, A. MARCINI, J.M. AKEY, R. DEKA, R.E. FEREL, M.W (1997) Ethnic-affiliation estimation by use of population-specific DNA. *American Journal of Human Genetics,* 60(3): 957–964.

SHUTLER, G., GAGNON, P., VERRET, G., KALYN, H., KORKOSH, S., JOHNSTON, E. & HALVERSON, J (1999) Removal of a PCR inhibitor and resolution of DNA STR types in mixed human-canine stains from a five year old case. *Journal of Forensic Sciences,* 44(3): 623-626.

SMITH, R. N (1995) Accurate size comparison of short tandem repeat alleles amplified by PCR. *Biotechniques, 18(1):* 122 – 128.

SMITH S., VIGILANT L., MORIN P.A (2002) The effects of sequence length and oligonucleotide mismatches on 5'-exonuclease assay efficiency. *Nucleic Acids Research,* 30(2): 111-113.

SOUTHERN, E. M (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology,* 98(3): 503-517.

SUDOYO, H., WIDODO, P.T., SURYADI, H., LIE, Y.S., SAFARI, D., WIDJAJANTO, A., KADARMO, D.A., HIDAYAT, S. and MARZUKI, S (2008) DNA analysis in perpetrator identification of terrorism-related disaster: Suicide bombing of the Australian Embassy in Jakarta 2004. *Forensic Science International: Genetics,* 2(3): 231-237.

SUTLOVIC, D., GOJANOVIC, M., ANDELINOVIC, S., GUGIC, D. & PRIMORAC, D (2005) Taq polymerase reverses inhibition of quantitative real time polymerase chain reaction by humic acid. *Croatian Medical Journal,* 46(4): 556-562.

SUNDQUIST, T (2010) Forensic Phenotyping: What a DNA can (and cannot) tell us about a criminal's appearance. *Forensic Science International: Genetics,* 45(3): 245-248.

TAKAHASHI, M., KATO, Y., MUKOYAMA, H., KANAYA, H. and KAMIYAMA, S (1997) Evaluation of five polymorphic microsatellite markers for typing DNA from decomposed human tissues - Correlation between the size of the alleles and that of the template DNA. *Forensic Science International,* 90(1): 1-9.

TAN, S., and DAVEY, C.A (2011) Nucleosome structural studies. S*tructural Biology,* 21(1): 128-136.

TEREBA, A (1999) Tools for Analysis of Population Statistics: *PowerStats. DNA, (*65)2: 14-16.

THANGARAJ, K., ET AL (2002) Is the amelogenin gene reliable for gender identification in forensic casework and prenatal diagnosis? *International Journal of Legal Medicine, 116* (2): 121 – 123.

THORISSON, G.A., SMITH, A.V., KRISHNAN, L. and STEIN, L.D (2005) The International HapMap Project web site. *Genome Research*, 15(11): 1592-1593.

THORISSON, G.A., and STEIN, L.D (2003) The SNP Consortium website: past, present and future. *Nucleic Acids Research*, 31(1): 124-127.

TONI, C., PRESCIUTTINI, S., SPINETTI, I. and DOMENICI, R (2003) Population data of four X-chromosome markers in Tuscany, and their use in a deficiency paternity case. *Forensic Science International,* 137(2): 215-216.

TONI, C., PRESCIUTTINI, S., SPINETTI, I., ROCCHI, A. and DOMENICI, R (2006) Usefulness of X-chromosome markers in resolving relationships: Report of a court case involving presumed half-sisters. *Forensic Science International, 247*(4): 414-416.

TAUTZ, D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research,* 17(3): 6463-6471.

TAUTZ, D (1993) Notes on definitions and nomenclature of tandemly repetitive DNA sequences, *DNA fingerprinting. Science, 345(3):* 21-28.

TREANGEN, T.J. and SALZBERG, S.L (2012) Repetitive DNA and next-generation sequencing: Computational challenges and solutions. *Nature Reviews Genetics,* 13(1): 36-46.

UTSUNO H, M. K (2004) Influence of template DNA degradation on the genotyping of SNPs and STR polymorphisms from forensic materials by PCR. *Bull Tokyo Dent Coll,* 45(4): 33-46.

VALLONE, P.M. and BUTLER, J.M (2004) Y-SNP typing of U.S. African American and Caucasian samples using allele-specific hybridization and primer extension. *Journal of Forensic Sciences,* 49(4): 723-732.

VALLONE, P.M., JUST, R.S., COBLE, M.D., BUTLER, J.M. and PARSONS, T.J (2004) A multiplex allele-specific primer extension assay for forensically informative SNPs distributed throughout the mitochondrial genome. *International Journal of Legal Medicine,* 118(3): 147-157.

VENTER, J.C (2001) The human genome. *Journal of Biolaw and Business,* 4(3): 98-102.

VOM STEIN, J. and HACHTEL, W (1988) Chloroplast DNA differences in the genus Oenothera subsection Munzia: a short direct repeat resembling the lambda chromosomal attachment site occurs as a deletion/insertion within an intron of an NADH-dehydrogenase gene. *Current Genetics,* 13(2): 191-197.

WALLACE, D.M (1987) Large- and small-scale phenol extractions. *Methods in Enzymology,* 152(1): 33-40.

WALSH, P.S., ERLICH, H.A. and HIGUCHI, R (1992) Preferential PCR amplification of alleles: mechanisms and solutions. *PCR methods and applications. Electrophoresis,* 1(4): 241-250.

WALSH, P.S., FILDES, N.J. and REYNOLDS, R (1996) Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. *Nucleic Acids Research,* 24(14): 2807-2812.

WALSH, S., LINDENBERGH, A., ZUNIGA, S.B., SIJEN, T., DE KNIJFF, P., KAYSER, M. and BALLANTYNE, K.N (2011) Developmental validation of the IrisPlex system: Determination of blue and brown iris colour for forensic intelligence. *Forensic Science International: Genetics,* 5(5): 464-471.

WALSH, S., LIU, F., BALLANTYNE, K.N., VAN OVEN, M., LAO, O. and KAYSER, M (2011) IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. *Forensic Science International: Genetics,* 5(3): 170-180.

WALSH, S., WOLLSTEIN, A., LIU, F., CHAKRAVARTHY, U., RAHU, M., SELAND, J.H., SOUBRANE, G., TOMAZZOLI, L., TOPOUZIS, F., VINGERLING, J.R., VIOQUE, J., FLETCHER, A.E., BALLANTYNE, K.N. and KAYSER, M (2012) DNA-based eye colour prediction across Europe with the IrisPlex system. *Forensic Science International: Genetics,* 6(3): 330-340.

WANG, D. Y., CHANG, C.-W., LAGACÉ, R. E., CALANDRO, L. M. & HENNESSY, L. K (2011) Developmental Validation of the AmpFℓSTR® Identifiler® Plus PCR Amplification Kit: An Established Multiplex Assay with Improved Performance. *Journal of Forensic Sciences,* 34(6): 845-855.

WANG, L., LV, M., ZAUMSEGEL, D., ZHANG, L., LIU, F., XIANG, J., LI, J., SCHNEIDER, P.M., LIANG, W. and ZHANG, L (2015) A comparative study of insertion/deletion polymorphisms applied among Southwest, South and Northwest Chinese populations using Investigator® DIPplex. *Forensic Science International: Genetics,* 21(1): 10-14.

WANG, Z., ZHANG, S., ZHAO, S., HU, Z., SUN, K. and LI, C (2014) Population genetics of 30 insertion–deletion polymorphisms in two Chinese populations using Qiagen Investigator® DIPplex kit. *Forensic Science International: Genetics,* 11(1): 12-14.

WATSON, R. & BLACKWELL, B (2000) Purification and characterization of a common soil component which inhibits the polymerase chain reaction. *Canadian Journal of Microbiology,* 46(7): 633-642.

WEBER, J.L., DAVID, D., HEIL, J., FAN, Y., ZHAO, C. and MARTH, G (2002) Human diallelic insertion/deletion polymorphisms. *American Journal of Human Genetics,* 71(4): 854-862.

WEI, Y., QIN, C., DONG, H., JIA, J. and LI, C (2013) A validation study of a multiplex INDEL assay for forensic use in four Chinese populations. *Forensic Science International: Genetics,* 9(0): 22-25.

WEIR, B.S. and COCKERHAM, C.C (1984) Estimating F-statistics for the analysis of population structure. *Evolution,* 38(6): 1358-1370.

WESTEN, A.S. MATAI, J.F.J. LAROS, H.C. MEILAND, M. JASPER, W.J.F. DE LEEUW, P. DE KNIJFF, T. SIJEN, A.A (2009) Tri-allelic SNP markers enable analysis of mixed and degraded DNA samples, *Forensic Science International: Genetics,* 55(3): 233–241.

WETTON, J.H., TSANG, K.W. and KHAN, H (2005) Inferring the population of origin of DNA evidence within the UK by allele-specific hybridization of Y-SNPs*. Forensic Science International,* 152(8): 45-53.

WHITAKER JP, C. T., URQUHART AJ, MILLICAN ES, DOWNES TJ, KIMPTON CP, GILL P (1995) Short tandem repeat typing of bodies from a mass disaster: high success rate and characteristic amplification patterns in highly degraded samples. *Biotechnique,* 18(2): 670-677.

WIEDBRAUK, D.L., WERNER, J.C. & DREVON, A.M (1995) Inhibition of PCR by Aqueous and Vitreous Fluids. *Journal of Clinical Microbiology,* 33(10): 2643-2646.

WIEGAND, P., & KLEIBER, M (2001) Less is more — Length reduction of STR amplicons using redesigned primers. *International Journal of Legal Medicine, 114*(4): 285 – 287.

WILLERSLEV, E., HANSEN, A.J. and POINAR, H.N (2004a) Isolation of nucleic acids and cultures from fossil ice and permafrost. *Trends in Ecology and Evolution,* 19(3): 141-147.

WILLERSLEV, E., HANSEN, A.J., RØNN, R., BRAND, T.B., BARNES, I., WIUF, C., GILICHINSKY, D., MITCHELL, D. and COOPER, A (2004b) Long-term persistence of bacterial DNA. *Current Biology,* 14(1): 9-10.

WILLUWEIT, S. ROEWER, L (2007) International forensic Y chromosome user group, Y chromosome haplotype reference database (YHRD): update. *Forensic Science International Genetics,* 44(1): 83–87.

WILSON IG (1997) Inhibition and facilitation of nucleic acid amplification. *Applied Environmental Microbiology,* 63(2): 3741-3751.

WONG, S.W., PABORSKY, L.R., FISHER, P.A., WANG, T.S. and KORN, D (1986) Structural and enzymological characterization of immunoaffinity-purified DNA polymerase and DNA primase complex from KB cells. *Journal of Biological Chemistry,* 261(17): 7958-7968.

WONG, Z., WILSON, V., PATEL, I., POVEY, S. and JEFFREYS, A.J (1987) Characterization of a panel of highly variable minisatellites cloned from human DNA. *Annals of Human Genetics,* 51(2): 4-6.

WRIGHT, S (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution,* 19(3): 395-420.

YANG D., ENG B., WAYE J.S., DUDAR J.C., SAUNDERS S.R (1998) Improved DNA extraction from ancient bones using silica based spin columns. *American Journal of Anthropology,* 105(4): 539–543.

YOUNG C, BURGHOFF RL, KEIN LG, MINAK-BERNERO V, LUTE JR, HINTON SM (1993) Polyvinylpyrrolidone-agarose gel electrophoresis purification of polymerase chain reaction-amplifiable DNA from soils. *Applied Environmental Microbiology,* 59(4): 1972-1974.

YOSHII T, TAMURA K, TANIGUCHI T, AKIYAMA K, ISHIYAMA I (1993) Water-soluble eumelanin as a PCR-inhibitor and a simple method for its removal. *Nippon Hoigaku Zasshi,* 47(3): 323-329.

ZAUMSEGEL, D., ROTHSCHILD, M.A. and SCHNEIDER, P.M (2013) A 21 marker insertion deletion polymorphism panel to study biogeographic ancestry. *Forensic Science International: Genetics,* 7(2): 305-312.

APPENDIX

Appendix 1: Ethical approval letter from the University of Central Lancashire's STEM Ethics Committee.

Appendix 2: *Locus-specific* information *of the Identifiler Plus Kit*

Appendix 3: Locus-specific information of the Investigator® DIPplex Kit.

Appendix 4: Allele frequency of 15 STRs in Life Technology™ Identifiler® Plus.

Appendix 1 : Ethical approval letter from the University of Central Lancashire's

STEM Ethics Committee.

4 December 2013

Will Goodwin / Majid Bashir
School of Forensic & Investigative Sciences
University of Central Lancashire

Dear William / Majid

Re: STEM Ethics Committee Application
Unique reference Number: STEM 155

The STEM ethics committee has granted approval of your proposal application '**Population studies of autosomal INDELS for different sub population groups in Qatar**'.

Please note that approval is granted up to the end of project date or for 5 years, whichever is the longer. This is on the assumption that the project does not significantly change in which case, you should check whether further ethical clearance is required.

We shall e-mail you a copy of the end-of-project report form to complete within a month of the anticipated date of project completion you specified on your application form. This should be completed, within 3 months, to complete the ethics governance procedures or, alternatively, an amended end-of-project date forwarded to roffice@uclan.ac.uk together with reason for the extension.

Yours sincerely

Kevin Butt
Vice Chair
STEM Ethics Committee

NB - Ethical approval is contingent on any health and safety checklists having been completed, and necessary approvals as a result of gained.

Appendix 2 : *Locus-specific* information *of the Identifiler Plus Kit*

Loci	*Location*	*Alleles*	*Dyes*	*9947A*
D8S1179	*8*	*8, 9 10, 11, 12, 13, 14, 15, 16, 17, 18, 19*	*FAM*	*13,13*
D21S11	*21q11.2-q21*	*24, 24.2, 25, 26, 27, 28, 28.2, 29, 29.2, 30,30.2, 31, 31.2, 32, 32.2, 33, 33.2, 34, 34.2, 35, 35.2, 36, 37, 38*	*FAM*	*30,30*
D7S820	*7q11.21-22*	*6, 7, 8, 9, 10, 11, 12, 13, 14, 15*	*FAM*	*10,11*
CSFIPO	*5q33.3-34*	*6, 7, 8, 9, 10, 11, 12, 13, 14, 15*	*FAM*	*10,12*
D3S1358	*3p*	*12, 13, 14, 15, 16, 17, 18, 19*	*VIC*	*14,15*
TH01	*11p15.5*	*4, 5, 6, 7, 8, 9, 9.3, 10, 11, 13.3*	*VIC*	*8,9.3*
D13S317	*13q22-31*	*8, 9, 10, 11, 12, 13, 14, 15*	*VIC*	*11,11*
D16S539	*16q24-qter*	*5, 8, 9, 10, 11, 12,13, 14, 15*	*VIC*	*11,12*
D2S1338	*2q35-37.1*	*15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25,26, 27, 28*	*VIC*	*19,23*
D19S433	*19q12-13.1*	*9, 10, 11, 12, 12.2, 13, 13.2, 14, 14.2, 15, 15.2, 16, 16.2, 17, 17.2*	*NED*	*14,15*
VWA	*12p12-pter*	*11,12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22,23,24*	*NED*	*17,18*
TPOX	*2p23-2per*	*6, 7, 8, 9, 10, 11, 12, 13*	*NED*	*8,8*
D18S51	*18q21.3*	*7, 9, 10, 10.2, 11, 12, 13, 13.2, 14, 14.2, 15,16, , 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27*	*NED*	*15,19*
Amelogenin XX,XY	*X: p22.1-22.3 Y: p11.2*	*X, Y*	*PET*	*XX*
D5S818	*5q21-31*	*7, 8, 9, 10, 11, 12, 13, 14, 15, 16*	*PET*	*11,11*
FGA	*4q28*	*17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 26.2,27, 28, 29, 30, 30.2, 31.2, 32.2, 33.2, 42.2,43.2, 44.2, 45.2, 46.2, 47.2, 48.2, 50.2, 51.2*	*PET*	*23,24*

Appendix 3 : Locus-specific information of the Investigator® DIPplex Kit.

DIP locus	localization	SNP ID	Motif (+DIP)
DIPplex Blue			
Amel. X	Xp22.1-22.3	M55418	TAAG
Amel. Y	Yp11.2	M55419	CACG
HLD77	7q31.1	rs1611048	TGGGCTTATT
HLD45	2q31.1	rs2307959	AGCA
HLD131	7q36.2	rs1611001	GCAGGACTGG
HLD70	6q16.1	rs2307652	GCACC
HLD6HLD111	16q1317p11.2	rs1610905rs1305047	CACA
HLD58	5q14.1	rs1610937	AGGA
HLD56	4q25	rs2308292	TAAGT
DIPplex Green			
HLD118	20p11.1	rs16438	CCCCA
HLD92	11q22.2	rs17174476	GTTT
HLD93	12q22	rs2307570	ACTTT
HLD99	14q23.1	rs2308163	TGAT
HLD88	9q22.32	rs8190570	CCACAAAGA
HLD101	15q26.1	rs2307433	GTAG
HLD67	5q33.2	rs1305056	CTACTGAC

DIPplex Yellow			
HLD83	8p22	rs2308072	AAGG
HLD114	17p13.	rs2307581	TCCTATTCTA
HLD48	32q11.2	rs28369942	CT CTGAAT
HLD124	22q12.3	rs6481	GACTT
HLD122	21q22.11	rs8178524	GTGGA
HLD125	22q11.23	rs16388	GAAGTCTGAGG
HLD64	5q12.3	rs1610935	ATTGCC
HLD81	7q21.3	rs17879936	GACAAA
			GTAAGCATTGT
DIPplex Red			
HLD136	22q13.1	rs16363	TGTTT
HLD133	3p22.1	rs2067235	CAACCTGGATT
HLD97	13q12.3	rs17238892	AGAGAAAGCTG AAG
HLD40	1p32.3	rs2307956	
HLD128	1q31.3	rs2307924	GGGACAGGTG G
HLD39	1p22.1	rs17878444	CCACTAGGAGA
HLD84	8q24.12	rs3081400	ATTAAATA
			CCTAAACAAAA ATGGGATCTTTC

Appendix 4 : Allele frequency of 15 STRs in Life Technology™ Identifiler® Plus.

D21S11					
	QATARI	PAKISTANI	SUDANESE	YEMENI	TUNISIAN
8	0.5%	0.5%	0.0%	0.5%	1.5%
9	1.0%	0.0%	1.5%	0.0%	0.0%
10	5.5%	12.5%	4.0%	8.0%	5.5%
11	9.0%	8.5%	9.0%	14.0%	10.0%
12	20.0%	12.5%	9.0%	19.5%	12.0%
13	20.5%	19.5%	25.5%	14.0%	30.5%
14	16.5%	19.0%	24.0%	19.5%	17.5%
15	18.0%	20.0%	19.5%	19.0%	17.5%
16	7.0%	7.0%	4.5%	4.5%	4.0%
17	1.5%	0.5%	2.0%	1.0%	1.5%
18	0.5%	0.0%	1.0%	0.0%	0.0%

CSF1PO					
	QATARI	PAKISTANI	SUDANESE	YEMENI	TUNISIAN
7	0.5%	0.0%	0.5%	0.0%	1.0%
8	1.0%	0.0%	3.5%	0.5%	3.0%
9	2.5%	1.5%	3.0%	4.5%	2.5%
10	31.0%	22.0%	34.0%	25.0%	26.0%
11	31.5%	31.0%	26.0%	34.0%	35.0%
12	27.5%	33.5%	29.0%	32.5%	28.5%
13	5.0%	10.0%	2.5%	3.5%	3.5%
14	1.0%	2.0%	1.5%	0.0%	0.5%

D7S820					
	QATARI	PAKISTANI	SUDANESE	YEMENI	TUNISIAN
26	0.0%	1.0%	0.0%	0.0%	0.5%
27	2.0%	2.0%	3.5%	3.5%	4.0%
28	16.0%	15.0%	13.5%	13.0%	7.0%
29	26.0%	16.5%	33.0%	24.5%	27.5%
30	22.5%	17.5%	21.5%	17.5%	26.0%
31	5.0%	5.5%	3.0%	6.5%	5.5%
32	0.5%	1.0%	2.0%	0.5%	1.5%
33	0.0%	0.0%	0.5%	0.0%	0.0%
34	0.0%	0.0%	0.5%	0.5%	0.5%
35	0.5%	0.0%	3.5%	1.0%	1.0%
36	0.0%	0.0%	2.5%	0.5%	0.0%
37	0.0%	0.0%	1.5%	0.0%	0.0%
25.3	0.0%	0.0%	0.5%	0.0%	0.0%
29.2	0.0%	1.0%	0.0%	0.0%	0.0%
30.2	0.5%	3.0%	1.0%	1.0%	1.0%
31.2	10.0%	15.5%	4.5%	6.5%	6.0%
32.2	10.0%	15.5%	5.5%	19.0%	10.5%
33.2	7.0%	6.0%	3.5%	5.0%	6.5%
34.2	0.0%	0.5%	0.0%	1.0%	2.5%

D3S1358					
	QATARI	PAKISTANI	SUDANESE	YEMENI	TUNISIAN
13	0.0%	0.0%	1.0%	0.0%	0.5%
14	5.5%	3.5%	5.0%	6.5%	5.5%
15	19.0%	31.5%	31.5%	19.5%	29.5%
16	32.5%	28.0%	28.0%	25.0%	27.5%
17	29.0%	22.5%	26.0%	33.0%	27.0%
18	13.5%	14.0%	8.0%	15.0%	9.5%
19	0.5%	0.5%	0.5%	0.5%	0.5%
15.2	0.0%	0.0%	0.0%	0.5%	0.0%

TH01					
	QATARI	PAKISTANI	SUDANESE	YEMENI	TUNISIAN
6	38.0%	31.5%	19.0%	28.0%	20.0%
7	15.5%	22.5%	40.5%	18.5%	21.5%
8	7.0%	17.0%	5.5%	14.0%	16.0%
9	24.5%	15.0%	27.0%	28.5%	27.0%
10	0.5%	1.5%	2.5%	0.0%	3.5%
11	0.0%	0.0%	0.0%	0.0%	0.0%
12	0.0%	0.0%	0.0%	0.0%	0.0%
13	0.0%	0.0%	0.0%	0.0%	0.0%
9.3	14.5%	12.5%	5.5%	11.0%	12.0%

D13S317					
	QATARI	PAKISTANI	SUDANESE	YEMENI	TUNISIAN
<8	0.0%	1.5%	0.0%	0.0%	0.0%
8	10.5%	16.5%	12.0%	20.5%	7.0%
9	5.0%	7.5%	6.5%	3.0%	4.0%
10	6.5%	9.0%	5.0%	6.0%	5.5%
11	27.5%	30.0%	26.0%	20.5%	28.0%
12	36.0%	25.0%	36.0%	35.0%	37.0%
13	12.5%	8.5%	7.5%	6.5%	13.5%
14	2.0%	2.0%	7.0%	8.5%	4.5%
>15	0.0%	0.0%	0.0%	0.0%	0.5%

D16S539					
	QATARI	PAKISTANI	SUDANESE	YEMENI	TUNISIAN
8	3.5%	7.5%	4.5%	3.0%	3.0%
9	12.5%	13.5%	16.5%	12.5%	13.0%
10	6.5%	5.0%	9.5%	7.0%	8.0%
11	40.5%	33.5%	31.0%	43.0%	35.5%
12	21.0%	31.0%	24.0%	20.5%	22.5%
13	13.5%	8.0%	12.5%	11.0%	16.0%
14	2.5%	1.5%	2.0%	3.0%	2.0%

D2S1338					
	QATARI	PAKISTANI	SUDANESE	YEMENI	TUNISIAN
15	0.0%	0.0%	0.0%	0.0%	0.5%
16	7.0%	1.5%	4.5%	8.0%	3.5%
17	20.5%	8.5%	19.0%	19.0%	27.5%
18	14.5%	14.0%	9.5%	12.5%	10.0%
19	11.5%	16.5%	19.5%	13.0%	9.5%
20	17.5%	12.0%	11.0%	15.0%	18.0%
21	6.5%	3.5%	5.5%	6.0%	7.0%
22	3.0%	5.5%	8.0%	4.5%	6.0%
23	6.5%	19.0%	9.5%	6.5%	8.0%
24	5.5%	12.0%	6.5%	11.5%	7.5%
25	5.0%	6.5%	7.0%	4.0%	2.5%
26	1.5%	1.0%	0.0%	0.0%	0.0%
27	1.0%	0.0%	0.0%	0.0%	0.0%

D19S433					
	QATARI	PAKISTANI	SUDANESE	YEMENI	TUNISIAN
9	0.0%	0.0%	0.5%	0.0%	0.5%
10	0.0%	0.0%	0.0%	0.0%	0.0%
11	1.5%	0.0%	2.0%	1.5%	0.0%
12	12.5%	7.0%	10.5%	10.0%	15.0%
13	21.0%	23.5%	24.5%	16.5%	28.5%
14	23.5%	29.5%	28.5%	25.0%	24.0%
15	15.0%	8.5%	10.5%	16.0%	18.0%
16	10.0%	8.0%	3.0%	3.0%	3.5%
17	0.5%	2.0%	1.0%	0.0%	0.5%
12.2	1.0%	0.5%	0.5%	0.0%	0.0%
13.2	3.0%	2.5%	5.5%	6.5%	1.0%
14.2	4.0%	4.5%	4.0%	7.5%	4.0%
15.2	4.5%	11.0%	6.0%	9.5%	3.5%
16.2	3.5%	3.0%	3.0%	4.0%	0.0%
17.2	0.0%	0.0%	0.5%	0.0%	1.5%
>17.2	0.0%	0.0%	0.0%	0.5%	0.0%

vWA					
	QATARI	PAKISTANI	SUDANESE	YEMENI	TUNISIAN
13	0.5%	1.5%	0.0%	0.0%	0.5%
14	6.0%	10.5%	14.0%	5.5%	14.0%
15	13.0%	10.5%	15.5%	10.5%	16.5%
16	24.0%	25.0%	26.5%	26.5%	24.5%
17	24.5%	23.0%	18.5%	27.5%	24.0%
18	21.5%	17.5%	17.0%	22.5%	15.0%
19	9.0%	10.0%	6.0%	7.0%	4.5%
20	1.0%	2.0%	2.0%	0.5%	0.5%
21	0.5%	0.0%	0.5%	0.0%	0.0%
22	0.0%	0.0%	0.0%	0.0%	0.5%

TPOX					
	QATARI	PAKISTANI	SUDANESE	YEMENI	TUNISIAN
6	1.0%	0.0%	0.5%	0.0%	1.5%
7	0.0%	0.0%	0.0%	0.0%	0.5%
8	50.0%	44.0%	44.0%	57.5%	42.5%
9	16.0%	11.0%	28.0%	9.5%	17.5%
10	9.0%	7.5%	11.0%	15.0%	11.0%
11	22.5%	33.5%	15.0%	17.5%	22.5%
12	1.5%	4.0%	1.5%	0.5%	4.5%

D5S818					
	QATARI	PAKISTANI	SUDANESE	YEMENI	TUNISIAN
8	1.5%	0.0%	13.5%	0.5%	2.0%
9	4.0%	4.5%	3.5%	7.5%	5.0%
10	13.0%	10.0%	9.0%	13.0%	7.0%
11	27.5%	31.0%	18.0%	22.5%	25.5%
12	33.5%	36.0%	34.5%	40.5%	39.5%
13	17.5%	17.5%	20.0%	14.5%	20.5%
14	3.0%	1.0%	1.5%	1.0%	0.5%
15	0.0%	0.0%	0.0%	0.5%	0.0%

	D18S51				
	QATARI	PAKISTANI	SUDANESE	YEMENI	TUNISIAN
9	0.5%	0.0%	0.5%	0.0%	0.0%
10	0.5%	0.0%	0.0%	0.0%	1.0%
11	1.5%	2.5%	1.5%	2.5%	1.5%
12	16.0%	11.0%	13.0%	16.0%	13.5%
13	22.0%	11.5%	9.5%	16.0%	13.5%
14	11.5%	24.0%	11.5%	21.0%	13.5%
15	12.0%	14.5%	13.0%	10.5%	8.5%
16	9.5%	13.5%	10.5%	10.5%	13.5%
17	8.5%	9.0%	13.5%	9.5%	13.5%
18	10.0%	5.5%	10.0%	9.5%	8.5%
19	4.0%	5.5%	5.5%	2.5%	6.5%
20	1.5%	1.5%	5.5%	1.0%	3.5%
21	0.0%	1.0%	1.0%	1.0%	2.5%
22	0.5%	0.0%	1.5%	0.0%	0.5%
23	0.5%	0.5%	0.5%	0.0%	0.0%
15.2	1.0%	0.0%	0.0%	0.0%	0.0%
16.2	0.5%	0.0%	0.5%	0.0%	0.0%
11.2	0.0%	0.0%	1.0%	0.0%	0.0%
10.2	0.0%	0.0%	0.5%	0.0%	0.0%
18.2	0.0%	0.0%	1.0%	0.0%	0.0%

FGA					
	QATARI	PAKISTANI	SUDANESE	YEMENI	TUNISIAN
17	0.5%	0.0%	0.0%	0.0%	0.0%
18	0.5%	0.0%	1.5%	0.5%	0.5%
19	4.5%	5.0%	6.0%	6.0%	6.0%
20	10.0%	8.0%	6.0%	7.5%	8.0%
21	8.0%	12.5%	8.5%	15.5%	16.0%
22	17.0%	20.0%	16.5%	13.0%	20.5%
23	17.0%	15.5%	18.0%	18.0%	17.0%
24	22.0%	16.5%	14.0%	21.5%	16.0%
25	11.0%	12.5%	12.0%	10.0%	11.0%
26	4.0%	4.5%	3.5%	5.5%	3.0%
27	1.0%	2.0%	4.0%	1.5%	1.0%
28	1.0%	0.0%	4.5%	0.5%	0.0%
29	0.5%	0.0%	4.5%	0.5%	0.5%
30	0.0%	0.0%	1.0%	0.0%	0.0%
18.2	0.5%	0.0%	0.0%	0.0%	0.0%
19.2	1.0%	0.0%	0.0%	0.0%	0.0%
20.2	0.0%	0.5%	0.0%	0.0%	0.0%
21.2	0.5%	1.0%	0.0%	0.0%	0.0%
22.2	1.0%	1.5%	0.0%	0.0%	0.5%
24.2	0.0%	0.5%	0.0%	0.0%	0.0%

D8S1179					
	QATARI	PAKISTANI	SUDANESE	YEMENI	TUNISIAN
7	0.5%	0.0%	0.5%	0.0%	1.0%
8	1.0%	0.0%	3.5%	0.5%	3.0%
9	2.5%	1.5%	3.0%	4.5%	2.5%
10	31.0%	22.0%	34.0%	25.0%	26.0%
11	31.5%	31.0%	26.0%	34.0%	35.0%
12	27.5%	33.5%	29.0%	32.5%	28.5%
13	5.0%	10.0%	2.5%	3.5%	3.5%
14	1.0%	2.0%	1.5%	0.0%	0.5%