Evaluation of Insertion/Deletion Polymorphisms (INDELs) Applied to Forensic Casework in Malaysia

by

Nur Haliza Hassan

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

May 2017

University of Central Lancashire

STUDENT DECLARATION FORM

Concurrent registration for two or more academic awards

Either *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

or ~~*I declare that while registered for the research degree, I was with the University's specific permission, a *registered candidate/*enrolled student for the following award:~~ N.

_____

Material submitted for another award

Either *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

or ~~*I declare that the following material contained in the thesis formed part of a submission for the award of~~ N.

_____

(state award and awarding body and list the material below):

* *delete as appropriate*

Collaboration

Where a candidate's research programme is part of a collaborative project, the thesis must indicate in addition clearly the candidate's individual contribution and the extent of the collaboration. Please state below:

Signature of Candidate _____

Type of Award PhD

School Forensic and Applied Sciences.

Those we love are always with us.
Their laughter and their wisdom,
their advice and thoughtfulness
are gifts of love that are ours to keep.

For mummy, Raja Azizah Raja Adnan and daddy, Hassan Lebai Mat.

ABSTRACT

In Malaysia, as well as other forensic laboratories in tropical climates many of the crime scene samples received at the forensic laboratory are less than ideal. They are often present low amounts and/or degraded due to environmental exposure to high temperatures, sun and humidity for days or even months. STR analysis is widely accepted by forensic community, but sometimes this technique gives unreliable results when profiling degraded samples as the amplicons size are relatively larger (100 bp to 450 bp). While, miniSTR is a reduced size of STR amplicons which enables higher recovery of information from degraded samples, but only a few loci are amplified and allele drop out still may occur, as the amplicons are up to 200 bp. The percentage of recovery DNA profile from degraded DNA using mtDNA is much higher due to its present in cells at a much higher copy number than the nuclear DNA. However, the major drawback for mtDNA is labour intensive and has a low information value (i.e. it is not highly discriminating).

Insertion/deletion polymorphisms (INDELs) are relatively new class of a DNA marker used in forensic casework; used most commonly as a supplementary method to STR (Short Tandem Repeat) based typing. INDELs, like SNPs (Single Nucleotide Polymorphisms), are particularly useful for the analysis of highly degraded DNA as the amplicon sizes are typically below 160 bp; they can also be valuable as an additional tool to help resolve kinship cases, with the advantage over STRs that do not have high mutation rates. INDELs have an advantage over SNPs in that they are length polymorphisms and so can be analysed by simply measuring the length of the allele(s). The Qiagen Investigator DIPplex® kit is currently one of two commercially available kits for the amplification of INDEL polymorphisms; it amplifies 30 biallellic INDEL loci and the amelogenin locus. The primers used are fluorescence labelled with 6-FAM, BTG, BTY and BTR. This technique is robust, relatively simple, and the results are analysed using the same capillary electrophoresis equipment and software as used for STR typing.

The INDEL markers have simple biallelic structure and combine the advantages of STR and SNP assays. This study has established that the INDEL technique, using the Investigator® DIPplex PCR kit, is a simple, informative and sensitive approach for the typing of degraded DNA, as compared to STRs and SNPs.

In this research, allele frequencies for 30 autosomal INDEL loci were studied in 500 unrelated individuals (100 each) from Malay, Malay-Chinese (M-Chinese), Malay-Indian (M-Indians), Iban and Bidayuh. The PCR amplification used the Qiagen Investigator® DIPplex kit. These population groups represent the majority of the population in Malaysia.

No significant departure from Hardy Weinberg Equilibrium (HWE) expectations were observed for most of the INDEL loci analyzed (*p*-value >0.05) on the Malaysian population samples. The exceptions were HLD101 for Malay ($p = 0.0009$), HLD133 for M-Indian ($p=0.005$), HLD125 for Iban ($p=0.028$) and HLD93 for Bidayuh ($p = 0.014$). However, when the Bonferroni correction for multiple testing performed on the population samples, none of the previous *p*-values was significant.

There were no Malaysian population studies was carried out using the Qiagen Investigator® Investigator DIPplex kit at the time of the research. This INDEL assay have undergo an extensive valdidation process and novelty report of the allele frequencies of INDELs would serve as reference database for individual identification in the Malaysian population in the future. Even the match probability of the STR is higher, INDEL still gives an acceptable value for forensic identification; e.g. linking different pieces of evidence or re-association of body parts in the case of human identification.

Biological samples received in Malaysia forensic laboratory have often been exposed to unfavourable environmental conditions. This can lead to DNA degradation and end up in incomplete DNA profiles. It is difficult to distinguish between low template DNA that are producing no or partial profiles because of DNA degradation and those that produce no or incomplete profile because of PCR inhibition. Even though real-time PCR methods are available for quantification and detection of PCR inhibitors, the information received is limited as real-time PCR targets amplicons that are much smaller than those typically targeted in forensic analysis. To gain more information on the quality of extracted DNA, a new multiplex PCR assay comprising a Mini 4-plex targeting amplicons of 50 base pairs (bp), 70 bp, 112 bp and 154 bp along with two Internal Amplification Controls (IACs) of 90 bp and 170 bp was developed. The primers were redesigned from a 4 plex & IACs system developed by previous PhD UCLan students. This multiplex was optimised so that it worked efficiently on DNA template as low as 0.009 ng, which highlighted the strength of the Mini 4-plex system. The IACs were effective in detecting PCR inhibitors. The Mini 4-plex system (Mini 4-plex & IACs) was demonstrated to be an effective tool for identifying degraded and inhibited samples, which could be used to triage forensic

samples in a casework laboratory. Therefore, this study has led to the improvement of new and novel markers assessing DNA degradation and PCR inhibition on forensic samples. This will demonstrate the compatibility with forensic laboratory workflows. The need of this Mini 4-plex assay in forensic laboratory can reduce time and cost of DNA analysis. Besides it will contribute to a good management samples, where after being assessed the samples can be decided to analyse using appropriate kit (e.g. miniFiler or INDEL). Indirectly, this will increase the quality of the sample itself.

In order to increase the power of a 15 Mini-INDEL multiplex, which was developed earlier by UCLan PhD student, a total of 9 autosomal INDEL markers that are not part of the Qiagen Investigator DIPplex® kit were selected and redesigned from (Pereira et al. 2009). In this study a simple and sensitive INDEL multiplex was successfully developed for human identification. However, the discrimination power is still low when compared to STR systems, but has potential value when analysing highly degraded material. By combining the 15 Mini-INDELs and 9 Mini-INDELs allele frequency data, it will give beneficial by increasing the match probability values in future analysis.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AGE	Agarose Gel Electrophoresis
AIM	Ancestral Informative Markers
BLAST	Basic Local Alignment Search Tool
CMP	Combined match probability
CNV	Copy number variation
CPD	Combined power of discrimination
CPE	Combined power of exclusion
ddNTP	Dideoxynucleotides
ds	double stranded
DNA	Deoxyribonucleic Acid
EA	East Asia
EDTA	Ethylenediaminetetraacetic acid
EXO	Exonuclease
GC	Guanine-cytosine content
HWE	Hardy-Weinberg Equilibrium
IAC	Internal amplification control
ICB	Intra-colour balance
IPC	Internal positive control
INDEL	Insertion/Deletion Polymorphism
ISEA	Island South East Asia
KIMIA	Department of Chemistry Malaysia
Kya	Thousand years ago
LOD	Level of detection
M-Chinese	Malay Chinese
MGS	Mamalian Genotyping Services

M-Indian	Malay Indian
MP	Match probability
mtDNA	Mitochondrial DNA
NCBI	National Centre for Biotechnology Information
NGS	Next generation sequencing
PCA	Principle component analysis
PCR	Polymerase chain reaction
PD	Power of discrimination
PE	Power of exclusion
PHR	Peak height ratio
PIC	Polymorphism information content
RAG	Recombination activating gene
RMP	Random match probability
SAP	Shrimp alkaline phosphate
SEA	South East Asia
SNP	Single Nucleotide Polymorphism
STR	Short Tandem Repeat
SWGDM	Scientific Working Group on DNA Analysis Method
TE	Tris-EDTA
TM	Melting temperature
XSTR	short tandem repeat (STR) on the X-chromosome
YSTR	short tandem repeat (STR) on the Y-chromosome

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

In the name of Allah, the Beneficent and the Merciful.

All studies performed in this thesis were conducted at the School of Forensic and Applied Sciences, University of Central Lancashire. This work was supported by the Department of Public Service Malaysia and Department of Chemistry Malaysia.

Scientific work is never a one-man show and I am grateful for all help, support and encouragement I have received throughout this 3 and half years. I would like to say TERIMA KASIH or THANK YOU to:

My supervisors, Dr William Goodwin and Dr. Arati Iyengar for sharing your deep knowledge within this field. Thank you for your guidance, encouragements and always having your doors open for discussions. Also, the FGG-team Dr. Hadi, Dr. Judith, Chrissy, Sharizah, Sasi, Rasheed, Temba, Majid, Tundey, Lais, Dina, Balnd, Hayder, and Ali for all your supports, big and small.

The Royal Police of Malaysia CSI team for valuable scientific collaborations and discussions, which I hope will continue.

My friends, especially ‘The Kampung Preston’ clans, Fit, Zaitul, Azni, Farah, Ieta, Rauhah, Eriza and Arni, thank you for keeping me sane and always being such good and supportive friends.

My family for your endless support and for believing in me during my journey.

Finally, my beloved husband, Hafiz and my children Khadeeja and Qayyum, thank you with all my heart for your constant support, love and caring. Your sacrifice meant a lot to me.

Thank you to all others who I did not mentioned but has been helpful in different ways.

CHAPTER 1

INTRODUCTION

The detection of deoxyribonucleic acid (DNA) polymorphism has over the years been used as a powerful tool in identification-tests, since its first debut use in a forensic casework investigation by Sir Alec Jeffreys in 1984 (Jeffreys 1985). Presently, the list of methods being used for DNA profiling continues to grow and are increasingly becoming more powerful, sensitive and rapid for human identification purposes.

Biological evidence submitted for forensic DNA analysis is often compromised due to degradation processes and/or the presence of inhibitors. The samples which are exposed to chemical, physical and biological insults, may in turn affect the success rate of DNA typing in different ways; environmental agents such as UV or radiation may affect the DNA structure while indirectly, the inhibitive effects of co-extracted agents such as humic acid or by-products of degradation may inhibit the PCR reaction. Such setbacks may therefore cause a loss of signal, peak height imbalances or allele dropout, which may result in investigators not being able to draw valid conclusions in relation to the implication of both suspects and victims in cases under investigation. The common short tandem repeats (STR) method used fails to produce the desired results with highly degraded samples owing to fragmentation of the DNA into small pieces.

This research can be divided into two components, which comprised:

a. To evaluate whether the Qiagen Investigator DIPplex kit is suitable for forensic application particularly for Malaysian casework samples. Especially for low template DNA samples which are exposed to tropical environment; high humidity and heat. Subsequently, to develop a new multiplex (mini INDEL) in order to analyse highly degraded samples.
b. to develop a new multiplex (Mini 4-plex & IACs) that can be used to assess the degradation and PCR inhibition of the DNA into small fragments (50 bp, 70 bp, 90 bp, 112 bp, 154 bp and 170 bp). This method able to triage samples prior to amplify to produce profiles; miniSTR or INDELs.

1.1 Forensic identity markers

Currently in Malaysia, protocols such as Short Tandem Repeats (STR), miniSTR, YSTR and mitochondrial DNA (mtDNA) are employed in forensic casework. STR analysis has widely been used and accepted by forensic community, but more often than not this technique reportedly produces incomplete results when the DNA is compromised or degraded as the amplicons' sizes are relatively larger (100 bp to 500 bp) (Butler 2007). However, the analysis of degraded DNA samples has been improved by using small-sized PCR products such as miniSTR (Butler, Shen & McCord 2003, Hughes-Stamm, Ashton & Van Daal 2011) and single nucleotide polymorphism (SNPs) (Martin et al. 2013, Martin et al. 2013). The success rate when profiling degraded DNA using mtDNA is much higher due to its presence in cells at a much higher copy number than nuclear DNA (1000 mitochondrial copies per cell). However, the major drawback for mtDNA is that it is laborious, complex interpretation and has a low information value (i.e. it is not highly discriminating) as it is a maternal lineage marker.

1.1.1 Short Tandem Repeats (STR)

Presently, STR typing is the most popular method for DNA profiling due to the highly polymorphic nature of STRs, robust amplification, sensitivity and high discrimination power. They consist of a short DNA sequence (typically 2-6 bp), arranged in tandem on particular region of a chromosome; they are also known as microsatellites. However, these assays do not always provide an excellent identification tool when it comes to degraded DNA which include their relatively large amplicon sizes 100 bp to 450 bp (Cotton et al., 2000, Krenke et al., 2002, Holt et al., 2002), 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 also make the analysis ambiguous especially in the case of mixture interpretation (LaRue et al., 2012) and their relatively high mutation rates (Weber et al., 2002). The commercial multiplexes produce amplicons approximately and in this case, allele dropouts are often observed at larger loci.

MiniSTRs were developed to recover information from samples which generated partial profiles (Butler, Shen & McCord 2003) but still the amplicon size is from 60 bp up to 280 bp (Butler 2011).These too, may not be applicable to caseworks with extremely degraded

DNA consisting fragments <100 bp in length (Zhang et al. 2009). In addition, STRs have higher mutation rates compared to other forensically relevant polymorphisms (Geada et al. 2003). Short amplicons binary markers (SABs) such as SNPs and insertion/deletion polymorphisms (INDELs) can be used as supplementary markers, which can add statistical value to cases that failed to be analysed using the STR method (Romanini et al. 2012).

1.1.2 Single Nucleotide Polymorphisms (SNPs)

SNPs analysis is another alternative method for forensic identification purposes especially for samples with low template or degraded DNA. SNPs are polymorphisms generated either by substitutions or insertion or deletion of a single nucleotide (Budowle, van Daal 2008). SNPs can be analysed using shorter amplicons than STR, they do not produce stutter during the PCR and their mutation rates are lower than STRs. SNP multiplexes can provide high power of discrimination, if sufficient polymorphic SNPs are assessed (Amorim, Pereira 2005, Sanchez et al. 2006). Gill et al. (2001) reported about 50 SNPs are required to have same discrimination power as 12 STRs. However, the analysis of SNPs can be laborious and time consuming (Garvin, Saitoh & Gharrett 2010), as it involves additional steps in comparison to STRs. For example, in mini-sequencing (or known as SNaPshot), two times of clean up steps were performed (refer Figure 1-1). Two different enzymes; exonuclease (EXO) and shrimp alkaline phosphate (SAP) were added to remove any excessive primers and dideoxynucleotides (ddNTPs) which will interfere the subsequent primer extension reaction. Then, the products were treated with SAP again to remove unincorporated fluorescent ddNTPs. If the SAP treatment incomplete, artefacts (e.g. dye blobs may occur. These, presents a barrier to the widespread adoption of these markers.

Since the completion of the human genome project in 2003, extraordinary progress has been made in genome sequencing technologies, which led to a reduction in the overall cost, maximizing the number and diversity of sequenced genomes and discriminatory power of applied DNA based assays (Goodwin et al., 2016). Next Generation Sequencing (NGS) was developed to enhance “first generation sequencing” technology, i.e. the Sanger method. Unfortunately, the practical utility of NGS in forensic genetics is yet unproven (Irwin et al., 2011, Bandelt, Salas 2012). Even though the advantages of the technology have been demonstrated on some difficult specimens (Briggs et al., 2009,

Krause et al., 2010, Scheible et al., 2014, Fordyce et al., 2015), the implementation on casework samples must be carefully weighed against their cost, in terms of effort, time (i.e. three days per run) and budget (Irwin et al., 2011). In addition, the experimental workflow and result interpretations are much more complex than the capillary electrophoresis (CE). Furthermore, neither NGS data quality assessments nor data analysis/interpretation procedures have yet performed according to the strict standard required for forensic genetics (Harismendy et al., 2009, Zaragoza et al., 2010). Other obvious limitations for forensic casework are the inability to detect longer STRs sequence, the need to use large amount and quality DNA in the ultra-sequencing platforms (Bandelt, Salas 2012).

1.1.3 Insertion/Deletion Polymorphism (INDELs)

To manipulate the small amplicons that can be generated for SNPs, and retain the relative ease of analysis offered by length polymorphisms at the same time, a new class of forensic DNA marker, called Insertion/Deletion Polymorphism (INDEL) (also known as Deletion Insertion Polymorphism (DIP) has been introduced by the forensic community.

This can be useful as a supplementary method to STR and miniSTR typing. INDEL markers are analysed using short amplicons, typically below 160 bp (e.g. 75 bp to 160 bp in the Qiagen Investigator DIPplex® kit), and therefore the chances are higher for success when analysing degraded samples. Furthermore, INDEL polymorphisms have relatively low mutation rates (largely considered to be unique events), are easier to interpret, and can be analysed using the same equipment and methods which are employed for STRs. In contrast, mtDNA has size of approximately 16,569 bp and the analysis include multiple procedures and complex interpretations.

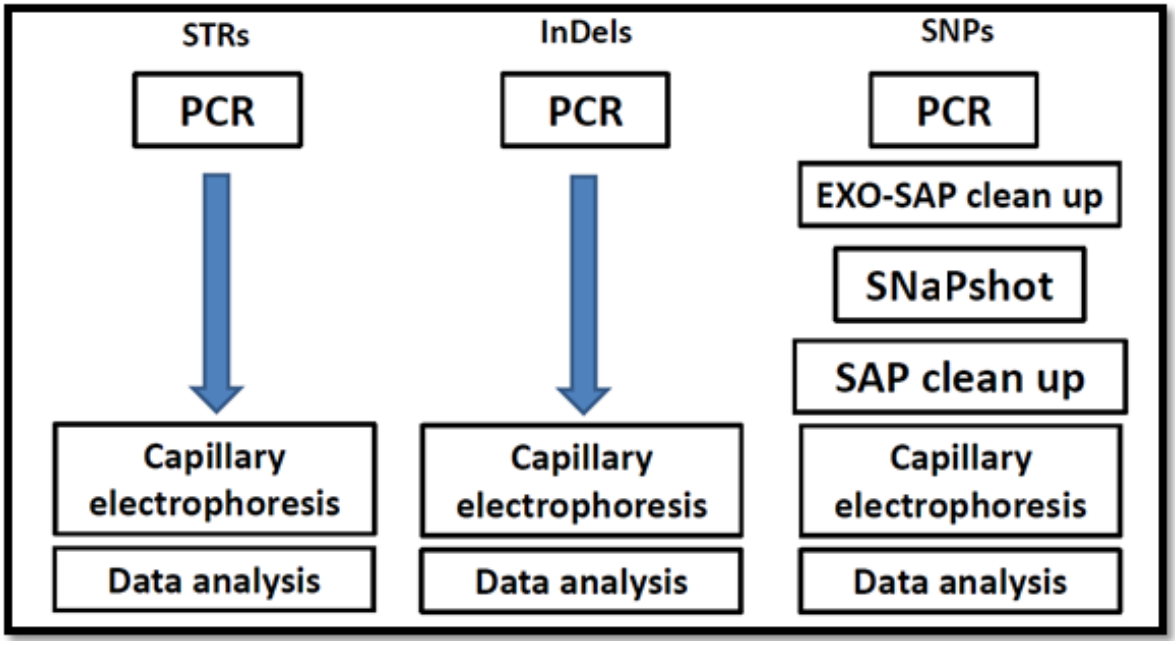


Figure 1-1: Comparison of methods among STRs, INDELs and SNPs. A straightforward method of INDEL detection is similar to STRs [adapted from Alvarez (2011)]. The INDELs can be analysed by means of multiplex with one PCR reaction and one CE (capillary electrophoresis) run; this approach is just the same as STRs. In comparison, SNP assays usually involve complex genotyping protocols and thus involving additional steps.

1.2 Discovery of INDEL primers

1.2.1 INDELs distribution and characteristics

Initially, the identification of INDELs in human genome were focused on human chromosome 22 and the re-sequencing data from 31 humans for this particular chromosome, only 13% of the genetic variations were identified as INDEL polymorphisms (Mullikin et al. 2000; Dawson *et al.* 2001).

INDELs have a number of characteristics that qualify them for forensic studies. Approximately 1.4 million human INDELs have been identified and characterized throughout the human genome (1000 Genomes Project Consortium 2012, Weber et al. 2002). The abundance of INDELs is important in differentiating one individual from another especially in human identity testing. The average number of INDELs per sample is 344,000 (1000 Genomes Project Consortium 2012) and the markers are very frequent

in human genome i.e., approximately 1 INDEL per 7.2 kb (Mills et al. 2006, Zidkova et al. 2013). In comparison, a map of 38 million of SNPs is distributed throughout the human genome, which providing an average of 3.60 million SNPs per sample (1000 Genomes Project Consortium 2012). It is noteworthy that approximately 3% (~ 93 kb) of the human genome is composed of microsatellites (Sawaya et al. 2013)**.**

Mutation rates among individuals are calculated based on the frequency of mutation arise and the expectations of it in an individual by referring to their parents. The mutation event is important for forensic DNA testing and the combination of the mutation is extremely important for genetic profile interpretations (Aşıcıoğlu, Oguz & Ozbek 2004) which are associated with the variations within a group of population. INDELs have low mutation rates, which are important when carrying out relationship testing and identification purposes (Pereira, Gusmao 2012). Both INDELs and SNPs have similar mutation rates in the range of 1 in 10^7 to 10^8 meiosis events (Amorim, Pereira 2005, Schneider 2012), whereas STRs have higher mutation rates (Geada et al. 2003, Neuvonen et al. 2012) i.e., 1 to 4 per thousand meiosis (Butler 2014). The lower mutational rate obtained, the more stable paternity and kinship analysis values are (Nachman, Crowell 2000, Phillips et al. 2008, Pontes et al. 2015). Therefore, the paternity index values can be improved using markers with a lower mutation rate (Pereira et al. 2012).

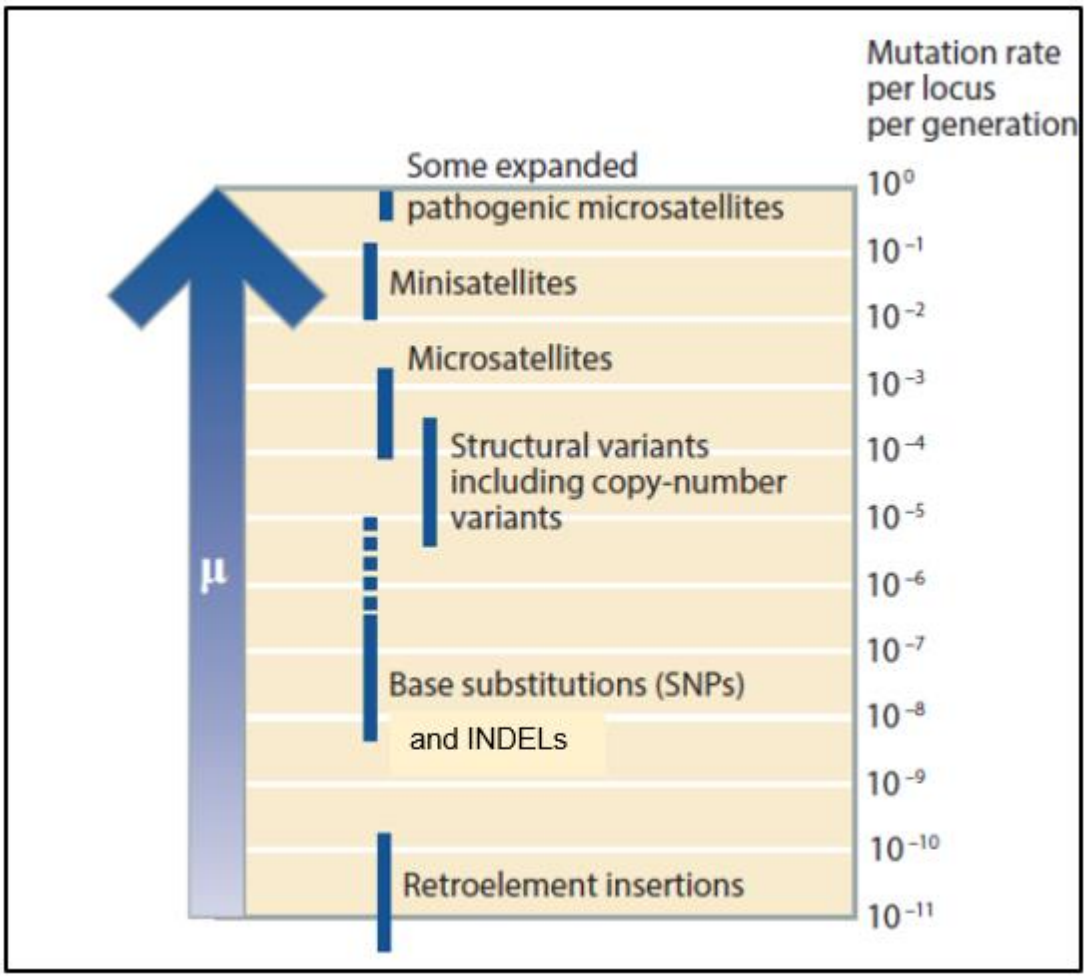


Figure 1-2**:** The mutation rates were edited from Jobling et al. (2014). INDEL's mutation rates similar to SNPs. Mutation rates of micro- and minisatellites, are variable between loci. The rates given here reflect ascertainment bias, in that they are those of the mostly widely used and polymorphic loci. However, the average rates are probably much lower.

According to (Mills et al. 2006), there are five classes of INDELs. These include: (1) insertion and deletion of single-base pairs, (2) monomeric base pair expansions (3) multi-base pair expansions of 2-15 bp repeat units, (4) transposon insertions and lastly (5) INDELs containing random DNA sequence between 2-9800 bp. An estimated 40% of INDELs belong to the last class; more than 99% of them are shorter than 100 bp (Mills et al. 2006, Zidkova et al. 2013).

1.2.2 The principle of INDEL

The INDELs have the simple structure of a biallelic marker; i.e. there are only two alleles per locus, based on insertion or deletions of fragments of DNA in the range of one to

hundreds of nucleotides; the two alleles can be categorized as either ‘long’ or ‘short’ (Butler 2005).

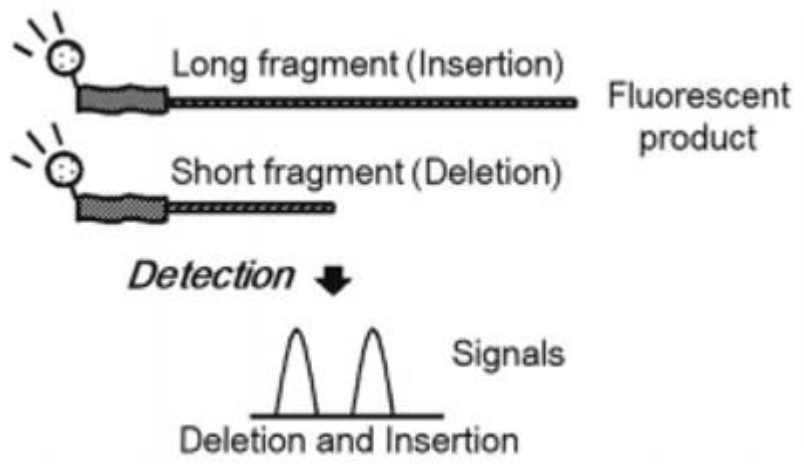


Figure 1-3: Detection of INDELs by universal fluorescent PCR. Two-step amplification produces the intermediate non-labelled fragments and the final fluorescently labelled fragments. The types of insertions and/or deletions are detected by fluorescent signals (edited from Oka et al. 2014).

1.2.3 INDELs database

The following section gives an historical account on how INDEL databases were generated and how INDEL markers have previously been selected from these databases for forensic purposes, such as the commercially available kit from Qiagen Investigator DIPplex kit. Prior to developing an INDEL multiplex, the initial candidate pool of markers were selected either from the previous work by (Weber et al. 2002) or online databases such as the National Centre of Biotechnology Information, dbSNP, the Marshfiled Clinic diallelic insertion/deletion database and Genome Browsers (e.g. UCSC Genome Browser, 1000 Genome Browsers and Ensmbl).

Overall, most of the markers selected were based on the following criteria:

i. biallelic autosomal INDELs
ii. within intron
iii. allele length variation 2 – 10 bp
iv. minor allele frequency > 0.1
v. mini allele frequency ≥ 0.25 in European, African and Asian population groups

vi. average heterozygosity ≥ 0.40

1.2.3.1 National Centre of Biotechnology Information (NCBI), dbSNP (Database of Single Nucleotide Polymorphism).

The NCBI Database of Single Nucleotide Polymorphism or commonly known as dbSNP, was officially changed in 2011 to database of Short Genetic Variations database (dbSGV) (Kitts et al. 2014). The database is a public- domain archive for a broad collection of simple genetic polymorphisms and it was developed to supplement GenBank (Benson et al. 1999, Benson et al. 2005). This collection of polymorphisms includes; SNPs, INDELs, retro-transposable element insertions and STRs (Sherry, Ward & Sirotkin 1999).

The SNP database can be accessed from the SNP homepage (http://www.ncbi.nlm.nih.gov/). To search and download the variations the links to the six basic dbSNP search options stated in the NCBI handout (ftp://ftp.ncbi.nih.gov/pub/factsheets/Factsheet_SNP.pdf) were used. The variations database also can be retrieved through the Entrez, which is a global query cross database search system (www.ncbi.nlm.nih.gov/Entrez/) features powerful options for constructing precise searches and managing results within and between the databases (Ostell 2014). By 2016, there were about 153 million human reference SNPs have been deposited into the public database dbSNP for build 147 (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_summary.cgi?view+summary=view+summary&build_id=147).

1.2.3.2 Marshfield Clinic

When Marshfield clinic was formed in 1916, the foundation was laid for a comprehensive medical system with patient care, research and education**.** This system has grown into one of the largest private, multispecialty group practices in the United States. As part of the history of Marshfield and Mammalian Genotyping Services (MGS), they have developed and maintained human linkage maps. These maps are available on the genetic maps page and all underlying information is available for scientist to develop markers, which can be found in the database or from other sources. Other markers such as SNPs and INDELs can be found on the INDEL marker search engine within the website.

They have characterized over 2,000 human INDELs (Weber et al. 2002). Supporting data for the (Weber et al. 2002) paper is located in the following two spreadsheets: MIDdata and MIDAlleleFreqs. The MIDAlleleFreqs table contains allele frequencies in five populations for 2000 of the Marshfield Insertion/Deletion Polymorphisms. The five populations included in this database are Europeans, Japanese, African Pygmies, Native Americans from the Amazon, and the Polymorphism Discovery Resource (PDR). Also included are the average values for all populations. The MIDdata table contains in addition to the allele frequency information, other information about the markers such as chromosomal location, inserted sequence, chimp/gorilla results, and data mining sources. Until September 2006, there are about 3.23 million genotypes available primarily to assist in linkage mapping of genes which cause or influence disease in humans and other model systems, under the direction of Dr. Jim Weber. The Mammalian Genotyping Service (MGS) have not been updated since they have completed a 12-year long contract in December, 2006.

1.2.3.3 UCSC Genome Browser

On 22nd June 2000, UCSC (University of California Santa Cruz) with collaboration with the International Human Genome Project consortium had successfully completed the draft of human genome assembly. UCSC then officially released the UCSC Genome Browser on website at http://genome.ucsc.edu/ (Kent et al. 2002).

1.2.3.4 The 1000 Genomes Project Consortium

The 1000 Genomes Project Consortium is the largest catalogue of human variation, including INDELs and genotype data by applying whole genome sequencing to a diverse set of individuals from multiple populations. In 2015, the 1000 Genomes Project Consortium have successfully completed a report on reconstruction of the genomes of 2,504 individuals from 26 populations (e.g. Africa, East Asia, Europe, South Asia and the Americas) via a combination of low coverage whole genome sequencing, deep exome sequencing, and dense microarray genotyping. The consortium discovered a total of 88 million variants that include 84.7 million SNPs, 3.6 million INDELs and 60 thousand structural variants. They reached their goal where >99% of SNPs with frequencies of at least 1% in the populations have been studied. Also the haplotype their LD patterns have been characterized (1000 Genomes Project Consortium 2012).

In order to make this project successful, the sample data for the project was collected in three phases. In Phase 1, about 1000 individuals were sequenced and phase 2 nearly 1600 individuals (included form Phase 1). Lastly, in Phase 3, the remaining individuals were sequenced. Whist, the pilot studies referred to different strategies of sequencing. The samples in each of the phases were sequenced using whole genome low coverage and full exome high coverage as well as genotyped on at least one high density genotyping platform. The data generated by the 1000 Genomes Project not only aided the interpretation of all genetic-association studies, but also provided lessons on how best to design and analyse sequencing-based studies of disease.

1.2.4.5 Ensembl

The Ensembl project (http://www.ensembl.org) generates genomic datasets through a system that is designed to analyse, store and distribute data and which enables interpretation through open data release (Yates et al. 2016). Like the UCSC Genome Browser, Ensembl also performs as a hub of reference and baseline data (Rosenbloom et al. 2015).

The Ensembl Variation database stores data imported from e.g. dbSNP and Sanger. There are different types of variants for several species including human, recorded as the following:

- single nucleotide polymorphisms (SNPs)
- short nucleotide insertions and/or deletions
- longer variants classified as structural variants (including CNVs)

The variants were classified into different categories and calculate the linkage disequilibrium for each variant. The variant was then grouped into different sets and specified into populations (e.g. African, American, East Asian, European and South Asian). In addition, allele frequencies and genotypes for each variant also available in the dataset which was imported from other projects (e.g. 1000 Genomes and HapMap). As of 2016, Ensembl had 157 million variant data for human from 6 sources e.g.; dbSNP, COSMIC, ClinVar, ESP, HGMP-Public and Phencode [http://www.ensembl.org/info/genome/variation/data_description.html#source].

1.2.4 Multiplex PCR amplification of INDEL loci

Multiplex polymerase chain reaction (PCR) is a variant of PCR in which multiple loci can be amplified by including more than one pair of primers in the same reaction (Markoulatos, Siafakas & Moncany 2002). It is cost efficient and less labour intensive than performing individual PCR as there is a decrease in the PCR reagents used and preparation time.

Since a multiplex PCR reaction is essentially multiple PCR reaction taking place at once, the success of the reaction lies in getting all the different components from each reaction to work together to produce successful amplification of all loci that make up the multiplex. On the other hand, not all primer selection criteria need to be met in order to synthesize a clean, specific product, since the adjustment of PCR conditions (e.g. composition of the reaction mixture, temperature, and duration of PCR steps) may considerably improve the reaction specificity (Rychlik 1993).

1.2.5 Commercial kits for the multiplex amplification of INDEL loci

Basically, the combination advantages of STRs and SNPs, abundance in human genome, lower mutation rates and short amplicons, are the characteristics in the development of INDEL assays for forensic purposes, especially for the profiling of degraded and inhibited samples (Romanini *et al*. 2012).

There are two autosomal INDEL kits that are commercially available for human identification testing (Larue et al. 2012); Qiagen Investigator DIPplex® and Mentype® DIPplex kits (Friis et al. 2012). Both use 30 biallelic INDEL markers and the sex marker, amelogenin in a single reaction.

In this research, the Investigator DIPplex® Kit (Qiagen) is used which detects using fluorescence-labelled primers using 6-FAM™, BTG, BTY, and BTR. The assay co-amplifies the repeat regions of the following 30 biallelic deletion/insertion polymorphisms (DIPs) loci HLD77, HLD45, HLD131, HLD70, HLD6, HLD111, HLD58, HLD56, HLD118, HLD92, HLD93, HLD99, HLD88, HLD101, HLD67, HLD83, HLD114, HLD48, HLD124, HLD122, HLD125, HLD64, HLD81, HLD136, HLD133, HLD97, HLD40, HLD128, HLD39 and HLD84 plus amelogenin gene in one reaction mix. These markers are disseminated among 19 autosomes and the distance of each one is at least 10 Mbp away from any

commercially available STR and SNP markers. The loci and the corresponding RefSNP (rs) numbers are presented in Table 1-1.

Based on the validation results on the Qiagen Investigator DIPplex® kit, ideally between 200 and 500 pg of DNA template should be used to obtain an optimum profile. Higher concentrations (above 1000 pg) may result in pull ups, over scaled peaks, strong signals and abnormal peaks (Pereira et al. 2012, Friis et al. 2012). In contrast, the INDELs assay has the potential to produce a reliable profile with < 100 pg of DNA.

In comparison, commercial kits may have advantages over the in-house INDEL multiplex; as they undergo extensive testing and quality control, ensuring that they can be used with confidence. Commercial typing kits are provided with allelic ladders that facilitate accurate typing. Kits are also highly sensitive, robust and needing very little DNA to obtain complete profile.

Table 1-1: The locus specific information of the 30 INDEL markers in Investigator DIPplex Kit (taken from (Investigator DIPplex Handbook 2010).

DIP locus	Chromosome localization	GenBank accession/SNP ID	Motif (+DIP)	Reference allele
Amel X	Xp22.1-22.3	M55418		
Amel Y	Yp11.2	M55419		
HLD77	7q31.1	rs1611048	TAAG	+DIP
HLD45	2q31.1	rs 2307959	CAGC	-DIP
HLD131	7q36.2	rs1611001	TGGGCTTATT	+DIP
HLD70	6q16.1	rs2307652	AGCA	-DIP
HLD6	16q13	rs1610905	GCAGGACTGGGCACC	-DIP
HLD111	17p11.2	rs1305047	CACA	-DIP
HLD58	5q14.1	rs1610937	AGGA	+DIP
HLD56	4q25	rs2308292	TAAGT	+DIP
HLD118	20p11.1	rs16438	CCCCA	-DIP
HLD92	11q22.2	rs17174476	GTTT	-DIP
HLD93	12q22	rs2307570	ACTTT	-DIP
HLD99	14q23.1	rs2308163	TGAT	-DIP
HLD88	9q22.32	rs8190570	CCACAAAGA	+DIP
HLD101	15q26.1	rs2307433	GTAG	-DIP
HLD67	5q33.2	rs1305056	CTACTGAC	-DIP
HLD83	8p22	rs2308072	AAGG	-DIP
HLD114	17913.3	rs2307581	TCCTATATTCTACTCTGAAT	-DIP
HLD48	2q11.2	rs28369942	GACTT	-DIP
HLD124	22q12.3	rs6481	GTGGA	-DIP
HLD122	21q22.11	rs8178524	GAAGTCTGAGG	-DIP
HLD125	22q11.23	rs16388	ATTGCC	-DIP
HLD64	5q12.3	rs1610935	GACAA	+DIP
HLD81	7q21.3	rs17879936	GTAAGCATTGT	-DIP
HLD136	22q13.1	rs16363	TGTTT	-DIP
HLD133	3p22.1	rs2067235	CAACCTGGATT	
HLD97	13q12.3	rs17238892	AGAGAAAGCTGAAG	-DIP
HLD40	1p32.3	rs2307956	GGGACAGGTGGCCACTAGGAGA	+DIP
HLD128	1q31.3	rs2307924	ATTAAATA	-DIP
HLD39	1p22.1	rs17878444	CCTAAACAAAAATG GGAT	-DIP
HLD84	8q24.12	rs3081400	CTTTC	-DIP

Ever since INDEL markers have been introduced to forensic communities, many research reports have been published and the findings lent credibility to the argument that INDELs cause genetic variation in human genes (Mullaney et al. 2010, Murthy et al. 2015). Therefore, the Diallelic INDEL markers are suitable in forensic practice and as ancestry informative markers (AIMS) to trace the origins of the populations; when the populations are not closely related.

Table 1-2: The comparison of power of discrimination between INDEL, STR and SNP on Caucasians (adapted from Investigator® DIPplex Handbook). The data support that the Investigator DIPplex® kit provides a powerful supplementary tool for human identification testing.

Kits	Loci	CPE/Trio[1]	CPI[2]	Population
Investigator DIPplex	30 DIPs	0.9980	2.83×10^{-13}	German Caucasian
AmpFISTR® Minifiler ™	8 STRs	0.99976	8.21×10^{-13}	US Caucasian
AmpFISTR® 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 (1)	52 SNPs	0.9998	5.00×10^{-21}	European Caucasian

[1]Combined probability of paternity exclusion. [2]Combined probability of identity

Table 1-3 The populations which have been evaluated using The Qiagen® Investigator DIPplex kit.

Continent	Populations	References
Asia	Koreans	(Kim et al., 2014, Seong et al., 2014)
	Chinese	(Wang et al., 2014, Shi et al., 2015, Meng et al., 2015, Zhang et al., 2015, Wang et al., 2016)
	Japanese	(Nunotani et al., 2015)
	Taiwan	(Pepinski et al., 2013)
	Bangladeshi	(Akhteruzzaman et al., 2013)
	Iran	(Poulsen et al., 2015)
Africa	South Africans	(Hefke et.al.,2015, Jacobs 2015)
	North Africans	(Moura-Neto et al., 2015)
	Somali	(Neuvonen et al., 2012)
Europe	Spain	(Martin et al., 2013)
	Portugal	(da Silva et al., 2013, Carvalho, Pinheiro 2013)
	Poland	(Pepinski et al., 2013)
	Greece	(Tomas et al., 2015)
	Finland	(Neuvonen et al., 2012)
	Czech	(Zidkova et al., 2013)
North America	Mexican	(Martínez-Cortés et al., 2015)
South America	Brazil	(Torres et al. 2014, Palha et al. 2015)
	Uruguay	(Saiz et al., 2014)

In addition, INDELs which proven for human identification have been characterized for population allele frequencies in the literature and publicly available databases, enable to be include into forensic statistical calculations (Gettings et al., 2015). At the time of this research was carried out, no investigation of INDEL markers using the Qiagen® Investigator DIPplex kit on the Malaysian population. Therefore, this application was a novel of this method where the Malay, M-Chinese, M-Indian, Iban and Bidayuh were successfully identified.

1.2.6 Non-commercial INDEL multiplex PCR

Reasons for some forensic laboratory develop in-house INDEL kit because the INDEL markers selected are specific to the selected population (i.e. China), ancestry studies and lastly to cut cost. The criteria of the non-commercial INDEL markers selected are

similar to the INDEL commercial kits markers, which described in Section 1.2.3. Several INDEL multiplex PCRs able to amplify up to 48 loci in one reaction have been established. There are multiplexes reported by using 20 (Huang et al. 2014), 21 (Zaumsegel, Rothschild & Schneider 2013), 29 (Li et al. 2011), 38 (Pereira et al. 2009, Oka et al. 2014), 40 (Pimenta, Pena 2010), 46 (Pereira et al. 2012) and 48 (Santos et al. 2010) biallelic autosomal INDELs. Besides for individual identification, these multiplexes were developed to predict the biographic ancestry of forensic casework based on insertion/deletion polymorphisms (Zaumsegel, Rothschild & Schneider 2013). Primers with low mutation rates were selected and loci that enabled differentiation of major ethnic populations (Pimenta, Pena 2010, Bastos-Rodrigues, Pimenta & Pena 2006) e.g. Europeans, Africans and Native Americans (da Costa Francez., P. A., Ribeiro-Rodrigues & dos Santos, S. E. B. 2012). In addition, 33 X-linked INDEL (X-INDEL) and Y-INDEL were developed as a complementary tool especially in complex paternity investigations (Freitas et al. 2010), (Pereira et al. 2011).

Table 1-4: The forensic parameters for non-commercial kits.

Number of INDEL markers	Population	CMP	CPE	References
20	South Central China	5.7 x 10-12	0.9989	(Huang et al., 2014)
29	Han Chinese	N/A	0.9930	(Li et al., 2011)
38	Africans (Angolans and Mozambicans),	2.9×10-13	N/A	(Pereira, Gusmao 2012)
	Europeans (Portuguese)	2.9×10-13	N/A	(Pereira, Gusmao 2012)
	Asians (Macanese and Taiwanese)	1.67 x 10-14	N/A	(Pereira, Gusmao 2012)
	Rio de Janeiro	3.2 x 10-15	N/A	(Manta et al., 2012)
	Japanese	2.12 x 10-15	N/A	(Oka et al., 2014)

1.2.7 INDELs: Potential applications and limitations

1.2.7.1 Human identification

The main use of this method is for the recovery of genetic information contained in challenged samples; degraded and inhibited ones (Pereira, Gusmao 2012, Klein et.al., 2015). This is important especially in attempting to link different pieces of evidence or re-associate of body parts in cases of human identification. INDELs are also suitable in cases of mixture interpretations due to the absence of micro-variants products (Larue et al., 2012). There are also Y-INDEL and X-INDEL; sex identification to enable efficient amplification from degraded samples, admixture population or for complex kinship analysis (Pereira et al. 2012, Martín et al. 2014).

1.2.7.2 Paternity and kinship studies

INDELs are the alternative method for paternity testing, especially in scenarios of a complex kinship investigation. In most cases, STR markers are able to provide powerful statistical results favouring one of the hypotheses due to high polymorphism (Amorim, Pereira 2005). However, there are also cases in which, the final LR values are insufficient (Ibarra et al., 2013). For instance, whenever the kinship is much more distant (Gomes et al., 2012) or the alleged father shows incompatible genotypes at a few loci along with a very high paternity index in the remaining systems (Pinto et al. 2013). In such cases, it is more likely that the alleged father could actually be a close relative of the individual concerned (son, father or brother) (Gomes et al. 2012, Karlsson et al. 2007), as in cases, of which, it is discovered that the alleged father and child are second-degree relatives (Pinto et al. 2013). In fact, the paternity index (PI) may rise sharply with the combination of STR and INDEL markers (Gao et al. 2015).

1.2.7.3 Ancestry Informative Markers (AIMs)

Ancestry informative markers or AIMs are polymorphic loci in which, the allele frequencies show distinct differences among populations (Investigator DIPplex Handbook 2010, Jobling, Gill 2004). AIMs are used to estimate the geographical origins of the ancestors of an individual, typical by continent of the origin of a DNA sample in which, the source of individual is unknown or the investigation is however unable to trace the ancestry (Phillips et al., 2007).

INDEL markers in AIMs have proven to be a valuable tool for estimating individual and global ancestry proportions within admixed populations as reported by (Santos et al., 2010) due to their stability, the density of distribution and a full range of allele frequencies pattern amongst the populations which are being looked into. These may also add up to the already existing valuable information by estimating the ethnic origin of crime samples without the presence of suspect(s).

1.2.7.4 Limitations

Apart from all the advantages discussed above, INDEL typing also has some limitations. INDEL markers are considered compatible as they can amplify small amplicon sizes. However, it should be noted that their biallelic property also implies that these markers are individually less-informative compared to microsatellites, which can be balanced by using a larger number of markers (Pimenta, Pena 2010, Santos, Pena & Pena 2015). Similar to SNPs, higher polymorphic multi-INDEL markers are required (Pimenta, Pena 2010, Santos, Pena & Pena 2015) to reach equivalent powers of discrimination or random match probabilities, to provide excellent discriminatory power which is either comparable to or exceeding that of STR markers (Rajeevan et al., 2012) which are widely used in forensic laboratories presently (Butler et al., 2007).

(Weber et al., 2002) reported, that not all INDEL markers are highly informative in all people, instead, some INDELs' alleles are relatively low in occurrence (minimum allele frequency) within some populations. This is so especially in populations that are closely related.

Caution should be exercised when working with low template DNA, as the providers usually recommended to use 32 cycles of amplification which produced unacceptable levels of artefacts (Bashir, Hassan 2016). Therefore, it can be difficult to detect contamination and result in a multitude of interpretations, worse still, a false interpretation (Larue et al. 2012, Murthy et al. 2015).

1.3 DNA degradation

There are a few factors that contribute to DNA degradation in forensic samples i.e. exposure to harsh environmental insults such as UV light, heat, water immersion and deterioration of tissues.

Malaysia has a high percentage of humidity (70% to 90%). Therefore, this could activate the activities of enzymes (i.e. endonuclease and exonuclease) which caused degradation. Endonuclease enzymes are enzymes that cleave to the bonds of the DNA from within the DNA molecule. However, the exonuclease enzymes that cleave to the nucleotides at the ends of the DNA molecule. Though the mechanism of degradation may differ, the end result is however universal; the fragmentation of a single or double helix transforming into smaller pieces as it degrades (Ballantyne 2006, Fondevila et al., 2008). This may make the amplification of DNA difficult and lead to low quality DNA or no DNA profiles (Kovatsi et al., 2009). Once the average DNA fragments are reduced to below 300 bp, a significant loss of genetic information may occur due to the lack of a suitable template DNA for amplification (Bender et al., 2004). Thus, the forensic samples need to be kept in dry place and cold temperature to deactivate the enzymes activities.

The degree of the biological degradation process depends on two factors: time and environmental conditions (Herrmann, Henke 1999). The processes may accumulate with the passage of time while environmental conditions (temperature, humidity, pH and soil chemistry) modifying the rate and aggressiveness of degradation (Fondevila et al. 2008). An increase in temperature and humidity may lead to growth of microbial and increased DNA activities (Dannemiller et.al., 2016). Both factors interact intricately so there is no direct correlation between time since death/deposition of material and the extent of degradation, making it difficult for troubleshooting of the samples.

As DNA degradation proceeds, three phenomena occur consecutively: loci imbalance, allele dropout and no amplification (see Figure 1-6). Allele dropout appears, as it increases along with the number of base pairs. This may result in partial profile. Allele may gradually get close to falling peak heights as the amplicon size increases. This phenomenon is commonly referred to as the degradation slope or ‘ski slope’ effect (Bright et al., 2013). The increasing detection of background contamination (pull-up, off scale peaks and incomplete adenylations) may lead to mistyping of DNA profiles (Kitayama et al., 2013).

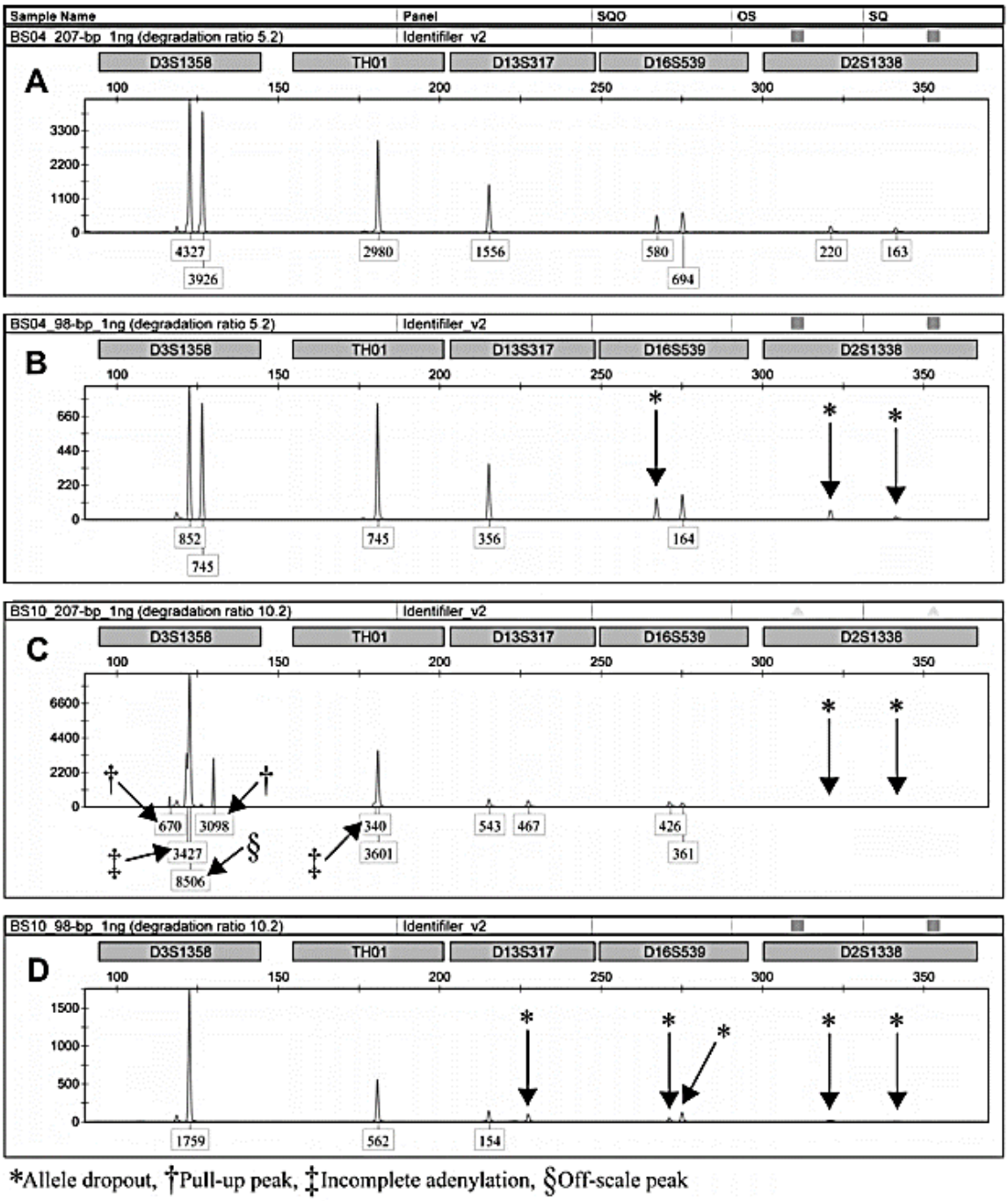


Figure 1-4: The panels represents a stepwise degradation of the same bloodstain sample (taken from Kitayama et al. (2013)). The y-axis represents the RFU while x-axis represents peak height. Different RFU scales y-axis are shown to demonstrate degradation. Panel A shows a reduction in fluorescence intensity, and a complete drop out of the larger STR. Allele dropouts were detected at larger loci as indicated in panel B, C and D. Plus, in panel C, pull up peaks and off-scale are detected in D3S1358 locus and incomplete adenylation was also detected in D3S1358 and TH01.

Degraded samples have been approached in possibly several different ways in order to estimate the level of DNA degradation among forensic samples. This includes the comet assays (Singh 2016), the development of species-specific PCR primers and agarose gel electrophoresis (Kaiser et al. 2008, Roy et al. 2016, Fredericks et al. 2012), species-specific primers and real-time PCR (Deagle, Eveson & Jarman 2006), Southern Blotting and hybridization (Alaeddini, Walsh & Abbas 2010), competitive PCR (Imaizuml, Miyasaka & Yoshino 2004) and profiling of DNA extracts with polymorphic marker systems (Dixon et al. 2006, Dong et al. 2016).

An in-house quadraplex was developed which can help detect the level of degradation and determine the length of DNA (e.g. 70 bp, 194 bp, 305 bp and 384 bp) prior to amplification (Nazir et al. 2013). The four primers pairs had been designed of the nuclear recombination activation gene 1 (RAG1) which involved in somatic (V(D)J) rearrangement of T and B cell lymphocytes. These primers are sensitive as it can amplify as low as 0.3 ng of DNA input. In addition, the 4-plex has potential to use as one of the screening tool for forensic analysis with degraded DNA by providing valuable information and minimize cost e.g the used of expensive kits.

1.4 PCR inhibition

The presence of inhibitory agents in a sample may also pose a significant challenge for PCR amplification. PCR inhibition, as reported in the literature, is the most common cause of amplification failures when sufficient amounts of DNA are present. Materials from the sample substrate may be co-extracted along with the target DNA dyes from denim (Schrader et al., 2012), tobacco from cigarettes (Hedman et al., 2010) and phenol chloroform (Schrader et al., 2012) or it may even be present within the sample itself e.g. calcium from bone (Thompson et al., 2014), heme in blood (Schrader et al., 2012) or melanin in hair (Thompson et al., 2014).

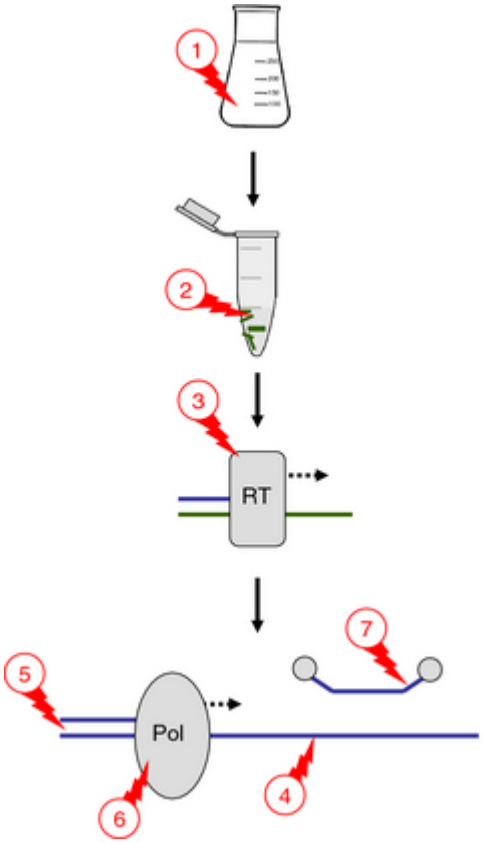


Figure 1-5: Schematic presentation of the attack points of polymerase chain reaction (PCR) inhibitors during sample preparation and PCR. The nucleic acids may interfere with surfaces of the vessels (1) or substances may react with nucleic acids (2) during sample processing and extraction. Other substances inhibit reverse transcription (3) and degrade or modify the template DNA (4). Annealing of primers to the template can be hampered (5) or the DNA polymerase is degraded, inhibited or altered (6). Finally, substances may interfere with binding of probes or with their fluorophores (7), adapted from Schrader et al. (2012).

According to (Kontanis et al., 2006), the mechanisms of PCR inhibition can be grouped into three categories based on the point of action during sample preparation and amplification (Wilson 1997). Inhibitors can interfere with cell lysis during DNA extraction (Jacobsen et al., 1992, Simon et al.,1996); degrade or capture nucleic acids (Lienert, Fowler 1992); or inhibit *Taq* DNA polymerase activity (Simon et al., 1996, Saulnier, Andremont 1992, Katcher, Schwartz 1994). While inhibitory mechanisms may vary, the outcome is general where lead to the loss of alleles or a complete failure of all loci. These results are similar to, and may be mistaken for highly degraded samples (Alaeddini et al., 2010). Table 1-4 shows some examples of matrices and their typical inhibitors.

Several techniques have been developed to detect PCR inhibitors. Quantitative real time PCR (qPCR) is a method of choice for quantifying the amount of DNA contained within

a sample. The PCR inhibitors can be detected by means of the alterations in either the efficiency of the reaction or by the changes observed in the threshold cycle (Ct), which indicates the lower concentrations of DNA are being amplified (Opel et.al, 2010). These real-time assays can simultaneously detect the presence of inhibitors by means of including an internal PCR control (IPC); internal amplification controls (IACs) also referred to as internal PCR controls (IPCs) (Zahra et al., 2011). The IAC is commonly a synthetic oligonucleotide that is co-amplified with the DNA samples. A reduction in the reaction efficiency, fluorescence values, and delay in amplification are indicators that certain samples may contain inhibitors or PCR failure. In contrast, the success of PCR can be indicated when the reference signal is produced with even no target DNA is present (Zahra et al., 2011). This assay is designed to be of general utility for forensic DNA quantifications purposes, but to be particularly useful for the post extraction analysis of samples that contain highly degraded DNA; especially samples discovered in instances of mass disasters, mass graves and missing persons (Swago K.L 2006).

PCR inhibition can be reduced during the extraction phase (Alaeddini 2012), however to do this without loss of phase DNA is a big challenge. Thus DNA extraction methods which favour the inhibitions elimination should be preferred e.g. PrepFiler™ Forensic DNA Extraction Kit (Applied Biosystems). This assay has been validated (Barbaro et al., 2009) and enables isolation of high quality DNA for PCR (Brevnov et al. 2009, Zimmermann et al., 2009, Pajnič et al., 2016) from biological samples (e.g. faeces, samples from leather and samples contaminated with soil) (Zimmermann et al., 2009). It uses polymer-embedded magnetic beads, which offer higher surface area (compared to larger magnetic bead systems) for capturing DNA molecules during the extraction process (Butler 2011, Barbaro et al., 2009). Other strategies to overcome PCR inhibitory effects includes the dilution of the extract to decrease the inhibitor. However, if the dilution factor was too high, it could prevent the success rate of DNA amplification (Marshall et al., 2013). This is because the DNA itself may equally be diluted. Therefore, it is better to get a purification kit instead (Kalmar et al., 2000, Bessetti 2007). An addition of *Taq* DNA polymerase or the use of more robust and tolerant DNA polymerases can also be used for troubleshooting purposes.

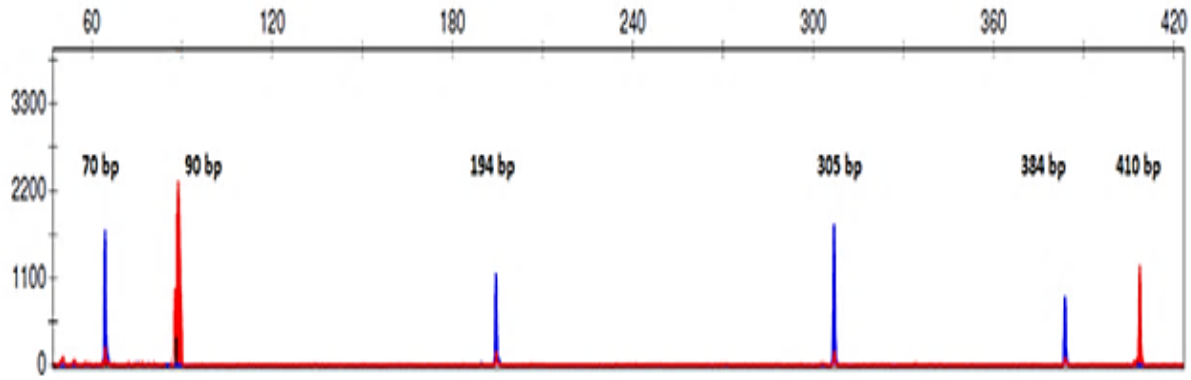


Figure 1-6: The electropherogram of the in-house multiplex system (4-plex and IACs) (Nazir et al. 2013) to asses and differentiate between degradation and inhibition of PCR in a DNA sample. The blue peaks are the 4-plex which to detect degradation, while the red peaks are the IACs, to detect inhibition.

Until today, there are still lack of in-house kits to assess the forensic samples, either it is degraded on inhibited. This research has led the re-developed and novel mini- 4-plex system as this technique could save more time and cost since less expensive reagents are involved while the 4-plex & IACs multiplex has to be analyzed on a Genetic Analyzer to evaluate the quality of a sample.

Table 1-5: An overview of PCR-inhibitory compounds and ion.

Compound or ion	Source	Mechanism(s)	Reference(s)
Al^3 ions	Sampling ion composition	Alters ion composition.	Wadowsky et al., 1994, Kobayashi et al., 2013
Bile salts	Faeces and plants	Direct effects on polymerase, sequestration and reduction in the availability of DNA template.	Thompson et.al.,2014, Monteiro et al., 1997, Lantz et al., 1997, Rådström et al., 2004, Al-Soud et al., 2005, Rouhibakhsh et al., 2008, Ghosh et al., 2009
Ca^{2+} ions	Bone	Competition with cofactor Mg^{2+}, alters ion composition.	Thompson et.al., 2014, Opel et.al., 2010, Rossen et al., 1992
EDTA	Anticoagulant	Chelating of Mg^{2+} and inhibit polymerase activities.	Schrader et al., 2012, Thompson et.al., 2014, Rossen et al., 1992
Free radicals	UV treatment of PCR tubes	React with polymerase.	Tamariz et al., 2006, Feine et al., 2016
Fulvic acid	Soil	Binds to polymerase.	Kreader 1996, Kasu, Shires 2015
Gum	Chewing gum	Effect on polymerase activities.	Hedman et al., 2010, Sanga et al., 2015
Haem/haematin and bilirubin	Blood	Release of iron ions, competition with template, lowers polymerase activities.	Schrader et al., 2012, Opel et.al., 2010, Rådström et al., 2004, Akane et al., 1994, Kermekchiev et al., 2009
Lactoferrin	Blood	Release of iron ions.	Al-Soud, Radstrom 2001, Geng, Mathies 2015
Myoglobin	Muscle tissue	Release of iron ions.	Belec et al., 1998, Fachmann et al., 2015
Phenol	Soil, chemical of DNA purification	Denaturation of polymerase, binding to polymerase.	Schrader et al., 2012, Schrader et al., 2012, Katcher, Schwartz 1994
Phytic acid	Faeces	Chelating of Mg^{2+} alters ion content if present as salt.	Fachmann et al., 2015, Thornton, Passen 2004
Polysaccharides	Faeces	Bind to polymerase.	Monteiro et al., 1997, Hu et al., 2015
Proteases (plasmin)	Milk	Degradation of polymerase.	Schrader et al., 2012, Wilson 1997, Powell et al., 1994
Tannic acid	Leather, plants, soil	Interaction with polymerase, chelating Mg^{2+} ions needed for the DNA polymerase.	Opel et.al., 2010, Kreader 1996, Geng, Mathies 2015
Tobacco	Moist snuff	Effect on polymerase activities.	Hedman et al., 2010, Sanga et al., 2015
Urea	Urine	Prevents non-covalent bonding, acts directly on polymerase and/or prevents primer annealing.	Geng, Mathies 2015, Khan et al., 1991, Mahony et al., 1998, Bergallo et al., 2006, El Bali et al., 2014

Bilirubin	Faeces	Form a complex with DNA polymerase, inhibit the Mg^{2+} ions and unbalance the ions in the PCR reaction by Ferric ions from heme.	Schrader et al., 2012, Kreader 1996, Kermekchiev et al., 2009, Al-Soud, Radstrom 2001
Cellulose	Cigarette filter paper, sampling filters, wood	Binds to DNA, effect on polymerase activities.	Hedman et al., 2010, Sanga et al., 2015, Bej et al., 1991, Lee, Cooper 1995, Watanabe et al., 2003, Hedman et al., 2009
Collagen	Bone	Interaction with *Taq* DNA, alteration of ion composition by binding cations.	Thompson et.al., 2014, Opel et.al., 2010, Scholz et.al., 1998, Kim et.al., 2000, Burkhart et.al., 2002
Ethanol	Chemical of DNA purification	Degradation of DNA polymerase.	Bessetti 2007, Rossen et al., 1992, Geng, Mathies 2015
Formaldehyde	Preservative	Interfere with DNA and DNA polymerase.	Geng, Mathies 2015, Johnson et al., 1995
Heparin	Anticoagulant	Binds to DNA, competition with template and/or interaction with polymerase.	Schrader et al., 2012, Al-Soud, Radstrom 2001, Satsangi et al., 1994
Immunoglobulin G	Blood	Formation of a complex with single-stranded DNA.	Fachmann et al., 2015, Abu Al-Soud 2000
Isopropanol	DNA extraction	Precipitation of DNA.	Rossen et al., 1992, Geng, Mathies 2015
Melanin	Skin, hair	Binds to DNA forms a reversible complex with polymerase.	Schrader et al., 2012, Thompson et.al., 2014, Opel et.al., 2010, Eckhart et al., 2000, Eckhart et al., 2000
SYBR green I	Detection dye	Binds to dsDNA with high affinity, binds to single stranded DNA (primers), affects polymerase activities.	Arezi et al., 2003, Zipper et al., 2004, Gudnason et al., 2007
Guanidinium	Guanidinium salts and Guanidinium thiocyanate for DNA isolation	Binds to the DNA molecules and effects the efficiency of *Taq* polymerase.	Thompson et.al., 2014
Humic acids	Soil	Quenches fluorescence, binds to DNA polymerase and to nucleic acid.	Kermekchiev et al., 2009, Tsai, Olson 1992a, Tsai, Olson 1992b, Watson, Blackwell 2000, Sutlovic et al., 2008
Indigo dye	Denim fabric and cigarette butts	Hinders qPCR detection through strong colouring interferes with SYBR Green I-DNA interactions.	Schrader et al., 2012, Opel et.al., 2010
Polymeric surfaces	Miniaturised real time PCR instruments	Binding of detection dye.	Gonzalez et al., 2007

1.5 Background of Malaysia and population genetic studies

1.5.1 The history of Malaysia

In 1400 AD, a prince from Palembang, Sumatera, with the name of Parameswara, founded Malacca after being defeated by the Kingdom of Majapahit. Under his rule, he built the Empire of Malacca Sultanate, which comprised most of the Peninsular and East coast of Sumatera. Due to its perfect location, Malacca became the most influential spice trade port in Southeast Asia and a meeting point for traders from East and West Asia. It is noteworthy that the migration at that time was not on a big scale, but intermarriages among them and with the Malays were common. Later, people from Java, Arabia, China and India started migrating to the Malay Peninsula.

The Malay Peninsula was colonized by the Portuguese (1511), Dutch (1641) and the British (1824). After the signing of Anglo-Dutch Treaty of 1824, the British brought in labourers from southern China to work in tin mines and those from southern India for rubber and tea plantations; this marks migration on a big scale to Malaya.

Sabah was invaded by the Sultanate of Brunei in the 17th century. Due to political disputes in Brunei, Sabah was ruled by two powers; the Sultanate of Brunei and Sultanate of Sulu. Sabah then was subsequently governed by the British North Borneo Chartered Company from 1881 until the Japanese occupation in 1942. After the World War II, Sabah became a British colony and accomplished independence from the British.

Sarawak was under the rule of Srivijaya kingdom from the 10th through the 13th century. The Malays of Sumatra then migrated and settled in the coastal areas of Sarawak and opened an entrepot port in Santubong, which was the most important port in the archipelago. After the fall of Srivijaya, Sarawak was taken over to be ruled by the kingdom of Majapahit and was subsequently taken over by the Sultanate of Brunei in the 19th century. In 1937, there was a commotion among the Brunei Rulers. James Brooke was asked to be the intermediary to bring the dispute to an end the rebellion. September 1841, James Brooke was appointed as the first governor of Sarawak. The Brooke dynasty (or known as White Rajahs) ruled Sarawak for 100 years.

The Federation of Malaya became an independent country on the 31st August 1957. On the 16th of September 1963, the federation expanded when states like Singapore, Sabah

and Sarawak were merged and the country was named as Malaysia. Since then, Malaysia has been known as a multi-racial country in Southeast Asia. However, Singapore opted to leave the federation on the 9th August 1965.

1.5.2 Demographics of Malaysia

Malaysia consists of two separate land areas, namely Peninsular Malaysia (West Malaysia) and East Malaysia. West and East Malaysia spread a total land area of about 330,290 km^2 and separated by approximately 40 miles of the sea (see Figure 1-8). There are eleven states and two federal territories in West Malaysia while East Malaysia is comprised of two large states and one federal territory. Malaysia essentially observes tropical weather, without extremely high temperatures; its typical daily temperature is in the range of 24 °C to 33 °C and throughout the year, the maximum temperature is in the range of 30 °C to 33°C on average (with humidity averaging between approximately 70% to 90%).

Figure 1-7: Map of Malaysia [adapted from Ezilon maps].

1.5.3 Population and ethnic groups in Malaysia

Based on the Department of Malaysian Statistics' portal, in 2015, Malaysia had a total population of about 31 million people from which, 24.6 million were found in Peninsular

Malaysia; 3.7 and 2.7 million respectively in Sabah and Sarawak. Currently, the annual growth rate is 1.9%; and it is estimated that by 2020, Malaysia's population may reach 32.49 million people (Khorasanizadeh et al., 2015).

The colourful mix of different ethnicity, cultures and religious makes Malaysian to be known as 'Malaysia, truly Asia'. Malaysia has more than 80 ethnic groups throughout the country (see Table 1-6) which divided into four main groups namely Bumiputera (including the Malays), M-Chinese, M-Indians and others (including non-Malaysian citizens) (Department of Statistics 2010, Department of Orang Asli Development Malaysia 2013). The great majority of Malaysians are made up of three major races comprising Malays, M-Chinese, and M-Indians in West Malaysia. The Peninsular Malaysia is also occupied by the indigenous communities, generally known as *Orang Asli* (or the native people). Orang *Asli* generally divided into three major groups, i.e. the Negrito, Senoi, and Proto-Malay. There are six sub-groups formed from each of the *Orang Asli* groups depending on their lifestyle and geographical areas. As East Malaysia consists of a wide kaleidoscope of ethnic groups living in Borneo; Sabah's population consists of 32 ethnic groups and the majority of them are indigenous Kadazan-Dusun. While the Sarawak population, consists of 40 ethnic groups and Iban is known to be the largest ethnic group. Figure 1-8 illustrates the distribution of ethnicity in Malaysia, the Malay (51%); Chinese (22%); M-Indian (6%); Bumiputera (12%); Non-citizens (8%) and others (1%).

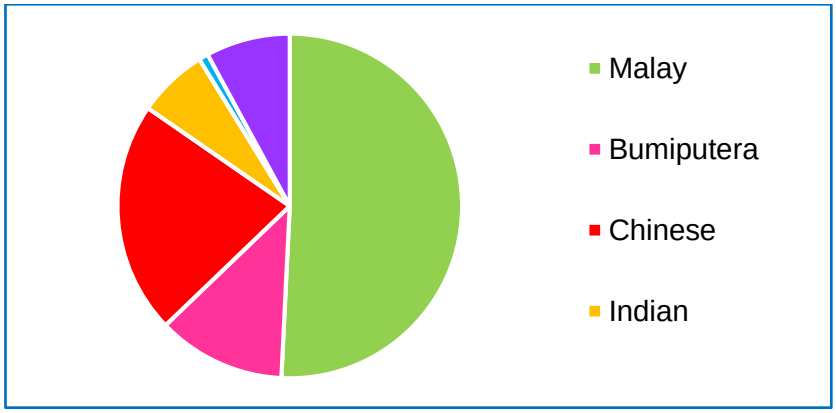


Figure 1-8**:** The proportion of Malaysian population.

Table 1-6: The classification of ethnic groups in Malaysia (Department of Statistics 2010).

BUMIPUTERA	THE CHINESE	THE M-INDIAN
Peninsular Malaysia Malay Negrito - Kensiu, Kintak, Lanoh, Jahai, Mandriq) Senoi - Temiar, Semai, Semoq Beri, Jahut, Mahmeri, Chewong Proto-Malay - Temuan, Semelai, Jakun, Orang Kanaq, Orang Kuala, Orang Seletar	Foochow Hainan Henghua Hokchia Hokchiu Hokkien Kantonis Hakka Kwongsai Teochew Other Chinese	Malabari Malayali Punjabi Sikh Sinhala Tamil India Tamil Sri Lanka Telugu Other M-Indians
BUMIPUTERA SABAH and W.P.LABUAN	**BUMIPUTERA SARAWAK**	
Melayu Melayu Brunei Kadazan Dusun Bajau Murut Balabak/Molbong Bisayah Bulongan Idahan Kadayan Lundayuh Orang Sungai Rungus Suluk Tidung Other Sabah Bumiputera	Melayu Iban Bidayuh Melanau Bisayah Sarawak Bukitan Iban Kadayan Sarawak Kanowit Kayan Kejaman Kalabit Kenyah Lahanan Lisum Lugat	Penan Murut Sarawak Punan Sabup Sekapan Sian Sipeng Tabun Tagal Tanjong Ukit Other Sarawak Bumiputera

1.5.4 Evolutionary history of DNA variability in Southeast Asia.

South Asia generally consists most parts of India, Pakistan and some other countries in the sub-Himalayan district, Myanmar, as well as Iran and Afghanistan. This region was among the first regions with modern human settlements 'Out of Africa' and it served as major route of dispersal to other geographical regions along the southern coastline route (Aghakhanian et al., 2015, Normile 2009), around 55 to 70 thousand years ago (kya) (Aghakhanian et al., 2015, Barker et al. 2007). In 2009, the HUGO Pan-Asian SNP Consortium research showed that there are two essential models, which clarified those initially inhabiting the continent. Firstly, there are two major waves of migration that was believed to have started from the Middle East; one followed a southern coastal route, starting from India and the journey continued from island to island across Indonesia, Malaysia and the Philippines to the Pacific; and secondly another wave carried over eastwards across the Eurasian steppe and curved south through the Asian mainland. In contrast a second theory suggested that that only one migration occurred; along the seaside course and the population moving towards north into East Asia from there according to this theory (Normile 2009).

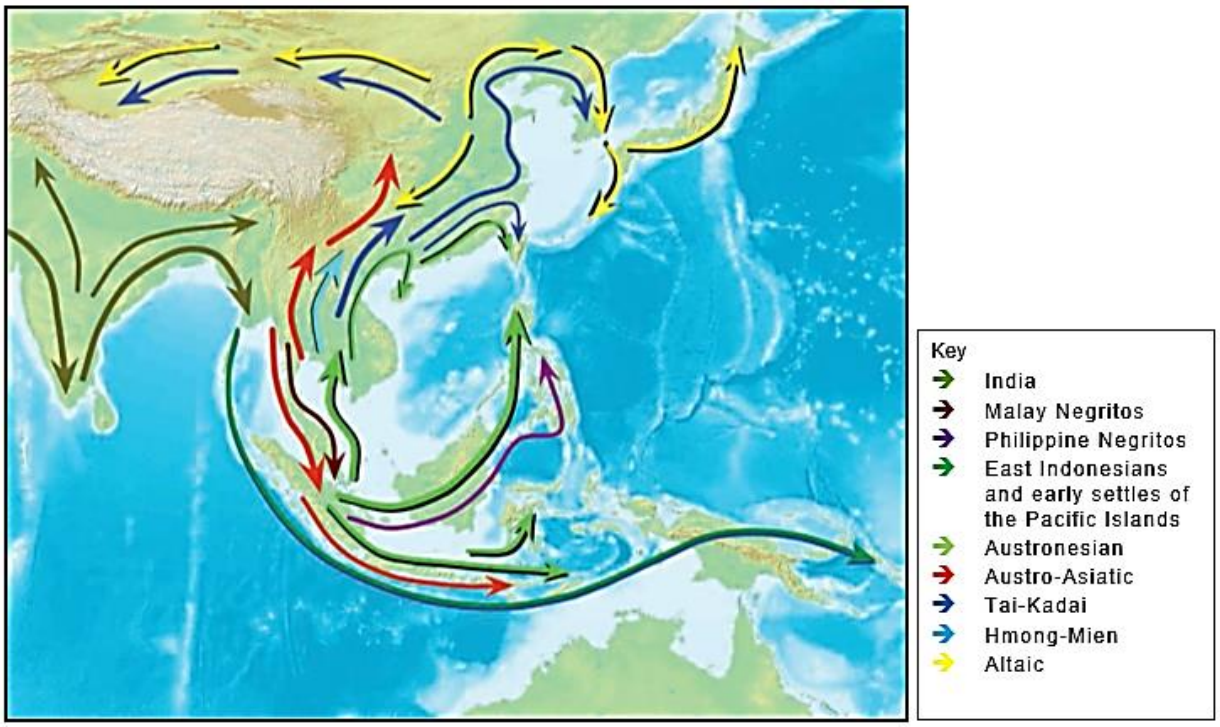


Figure 1-9: Genetic diversification of humans after migration from South Asia coast and split into numerous genetically distinct groups that moved across Southeast also migrated north into East Asia (Normile 2009).

The above figure demonstrates possible courses of pre-history relocation of human populations based in Asia. From their research, it was discovered that the modern mutual ancestors of Asians arrived first in India (moss green). Later, some of them travelled to Thailand and southwards to the lands that are today known as Malaysia, Indonesia, and the Philippines. The earliest immigrants must have travelled very far south before they settled down in one place. These involved the Malay Negritos (brown), Philippine Negritos (purple), the East Indonesians, and early settlers of the Pacific Islands (green). Then, some of the groups migrated towards North, which culminated in the admixture with the earlier inhabitants existed there and, lastly, formed different populations which we now refer to as Austronesian (lime green), Austro-Asiatic (red), Tai-Kadai (dark blue), Hmong-Mien (light blue), and Altaic (yellow).

1.5.5 Genetic studies on Malaysian population

Since 1960s, Malaysia has been a focal centre of interest of the human population geneticists due to the variety of ethnic groups (Tan 2001). Research on autosomal DNA, SNPs, mtDNA, Y-chromosome, blood groups, human platelet antigen, human leukocyte antigen, human neutrophil antigen and killer-cell immunoglobin-like-receptor have been carried out to support the complex genetic history of Peninsular Malaysia (Norhalifah et al., 2016).

1.5.5.1 Orang Asli, Indigenous people in Sabah and Sarawak

The Orang Asli are the earliest settlers who existed in Peninsular Malaysia (Kamarudin, Ngah 2007). Usually, they were found in smaller groups and lived scattered all over the Peninsula. *Orang Asli* are classified into three main groups i.e., Semang, Senoi, and Proto- Malay (aboriginal Malay) based on the linguistic, physical and anthropological characteristics (Aghakhanian et al., 2015). The Semang represent the first human settlers In Peninsular Malaysia since 50 kya (Norhalifah et al., 2016). While the Senoi migrated from Indochina, following the migration route through the mainland Southeast Asia before settling in the Malay Peninsula and the Indonesian archipelago (Kamarudin, Ngah 2007). However, the Proto-Malays migrated last to this region and with some ‘Negritos’ admixture. This culminated in the emergence of the contemporary Malays (Norhalifah et al., 2016). Their features are similar to each other; the Semang ‘Negritos’ have dark skin complexion and woolly hair, the Senoi who are distinguished by the wavy hair while the Proto have tan skin and straight to slightly wavy hair (Bulbeck 2011).

The Sabahan indigenous people bear more similarities with the mainland South East Asia/East Asia (SEA/EA) group than the Island of South East Asia (ISEA); as the Kadazan-Dusun, Bajau and Rungus clustered together with the Han Chinese, Japanese, Cambodian and Vietnamese (Kee et al., 2012).

Studies showed the Iban gave high genetic affinity to mainland SEA populations (Simonson et al. 2011). Therefore, the native populations of East Malaysia could be the ancestors of early settlers from the mainland of SEA or EA. (Chang et al., 2009) reported, that the Y-STR profiles showed that the three indigenous (Ibans, Bidayuhs, and Melanaus) are distinctly different from each other, also from the Malays, Chinese, and M-Indians.

1.5.5.2. The Malay, M-Chinese and M-Indian

The Malay population is made up of a mixture of a group of people which is still extant in South East Asia (SEA) as early as 3000 years ago (Andaya 2001). Thus, the Malays are genetically related to the *Orang Asli* (particularly Semang and Senoi) regardless of differences in relation to their physical features (Deng et al., 2014). However, specifically the Chinese, the M-Indians, the Arabs, and Europeans, have significantly affected the region, and have substantial impact in modern centuries due to colonization and globalization of the whole human culture, thus culminating in the mosaic genomic pattern in the Malays (Deng et al., 2014, Deng et al., 2015).

The Malaysians Chinese ancestry are descendants to the migrants from Southern China, originating from the Southern Mongoloid ethnic groups of Southeast Asia (Tan 2001), and Malaysian-Indians who are of South-Indian ancestry.

1.6 Forensic laboratory in Malaysia

In 1996, DNA profiling was first used in Malaysia, pioneered by the DNA Forensic Laboratory, Chemistry Department of Malaysia. The laboratory carried out DNA analysis on evidentiary biological samples such as bloodstains, seminal stains, hair and other biological samples discovered in the investigation of murder, suspicious deaths, body identification, rape and other sexual offences, assault, and also paternity cases. There has been a sharp rise in the number of cases involving DNA profiling ever since.

Yet, like other forensic laboratories in tropical climates, many of the crime samples received are less than ideal. They are often available in low amounts and/or degraded due to the factors like environmental exposure to heat, sun, and humidity for days or even months.

Additionally, the presence of inhibitors, bacteria and mould can affect the quality of DNA, which in turn may result in incomplete DNA profiles due to allelic or locus drop out, thus reducing the evidential value of the forensic evidence which is required to be produced in courts of law. Sometimes no DNA profiles are generated from the degraded samples.

Other main issues affecting the success of DNA typing are the time gaps between sample collection stages and the submission of the exhibits to KIMIA, along with packaging techniques and storage facilities. Figure 1-10 shows the flow chart of collection of biological evidence from crime scenes by the CSI (Crime Scene Investigators) until submission for DNA analysis.

Currently in Malaysia, protocols such as STRs, miniSTRs, YSTRs and mtDNA used for DNA profiling.

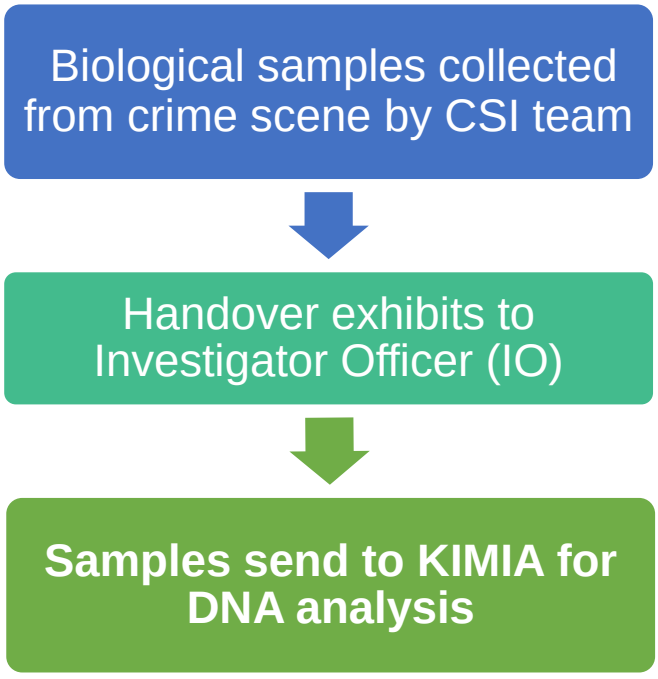


Figure 1-10: Flow chart of evidence collection for DNA analysis in Malaysia.

1.7 Overview and aims of the study

Over the years, scientists have developed new techniques they believe could analyse degraded DNA. Developing new scientific techniques will not only grasp the attention of other researchers in the forensic community but will show society how technology has improved throughout centuries. Through active research forensic scientist are improving issues within the criminal judicial system from a comprehensive perspective. The aim of this study is to improve the efficiency of INDELs for forensic identification and typing challenging of highly degraded samples. To achieve this, following objectives have been identified:

Chapter 3

a. To optimise and validate the INDEL system for processing of crime scene, bone and reference samples.
b. To determine the INDEL typing success of problematic specimens. Several casework and bone samples were evaluated using the Qiagen Investigator DIPplex kit in order to observe the sensitivity of the assays in generating profiles from degraded DNA samples. These profiles were compared with the standard STR typing which has been carried out on the same casework and bone samples.
c. To evaluate by investigating the sensitivity, reproducibility and robustness of the kit.
d. The 30 loci in the Qiagen Investigator DIPplex kit were evaluated in five Malaysian populations, i.e the Malays, M-Chinese, M-Indian. Iban and Bidayuh.
e. To optimise and validate the INDEL system for processing of crime scene, bone and reference samples

Chapter 4

a. Evaluation of the population genetic data of 30 autosomal INDELS (using Qiagen Investigator DIPplex Kit) in the five Malaysian subpopulation groups (approximately 100 samples each). Standard population biology parameters, such as F_{ST} were calculated and also through simulation exercises using software such as Arlequin to gain a picture of INDEL polymorphisms, and how they differ in each ethnic groups.
b. To estimate the effectiveness of INDELs for crime scene work and kinship analysis. This will be done by calculating match probabilities derived from the

population samples and also non-probative crime scene and bone samples processed using this technology. This data will be compared with the standard STR markers and SNPs, which have been carried out on Malaysian population.

c. Identifying, optimization and characterization the Malaysian ethnic groups of ancestry informative markers sets (AIMS) in order to provide a resource for assessing continental ancestry in a wide variety of genetic studies, particularly in INDELs. The Snipper app suite v2.0, developed by the University of Santiago de Compostela was used in this research in order to assess the usefulness of Qiagen Investigator® DIPplex kit among the Malaysian population in attempting to trace the ancestral origin of a DNA sample. The classification success could be estimated by using the cross validation option. The ability of this algorithm method for forensic ancestry inference varies significantly from one population to another (Yun et al. 2015). But, if biogeographic ancestries are very close, results may not be predicted accurately using SNIPPER (Jiang et al. 2015)..

Chapter 5

a. Develop a new PCR assay, Mini 4-plex by adding four new markers (50 bp, 70 bp, 112 bp, 154 bp) to assess the degree of degradation and to study the fragments that has ranges between 50 bp to 150 bp.
b. To combine Internal Amplification Controls (IACs) with a min 4-plex PCR assay to differentiate degraded and inhibited samples. Attempts were made to combine an existing mini 4-plex PCR assay with two newly designed IACs; IAC_{90} and IAC_{170} and thus develop a new multiplex PCR assay. This new multiplex was validated to study its sensitivity and capability in amplifying DNA of various concentrations.
c. Assess the new PCR assay on different types of forensic biological samples. The optimised assay was assessed for it efficiency in identifying the degraded and inhibited biological samples.

Chapter 6

a. Use the new PCR assay for quantification of DNA. The PCR assay was used to amplify various concentrations of control DNA and by their peaks heights, plotted into a correlation graph for the purpose of quantification.

Chapter 7

c. To develop a multiplex of mini INDELs, based on the results of the commercial kit with highly degraded samples, and its application on a set of Malaysian population samples. Efforts were taken to combine the existing Mini INDEL primers with Majid Bahir's (UCLan PhD student) in-house Mini INDELs and therefore develop a new multiplex PCR assay. An allelic ladder was also developed for the multiplex.

CHAPTER 2

METHODS AND MATERIALS

2.1 Overview

The methods and materials described in this chapter have been applied in all the studies detailed in this thesis. Any specific methods which are applied to certain analyses have been described in the relevant chapters.

Good laboratory practices were adopted while doing this research. Clean lab coats and disposable gloves were worn during any laboratory work to reduce contamination. Each batch of extractions was carried out together with a reagent blank to identify any contamination during extraction process. Prior to any experiment, the bench top surface and the equipment were cleaned with ethanol. Sample preparation for PCR was carried out using dedicated pipettes in a PCR hood together with positive and negative controls. The positive controls used in this research were supplied with the kits, whilst the negative control was distilled water. All samples in each experiment were prepared in duplicate and the final volume of the extracted samples was fixed at 40 µl unless stated otherwise. This allowed for easier comparisons between methods.

2.2 Health, safety and ethics

Risk assessments and COSHH training have been completed before the use of instruments and chemicals in any experiment (refer to Appendix II). Ethical approval was obtained from the University of Central Lancashire's Health, Safety, and Ethics Committee prior to the experimental work (refer to Appendix III).

The permission to collect forensic samples from Malaysia was obtained from the Forensic Division, Department of Chemistry Malaysia (refer to Appendix IV**)**.

2.3 Preparation of samples

2.3.1 The Malaysian population samples

The samples comprised approximately 1 ml blood spotted onto 100 FTA® cards (Whatman® Bioscience). The FTA card contained chemicals that lyse cells, denature proteins, and protect nucleic acids from nucleases. The bloodstained specimens were for population studies from unrelated individuals of each of five ethnic groups (the Malays, M-Chinese, M-Indians, Iban and Bidayuh), which were collected with informed consent. From the particulars, the FTA cards were labelled with their names and ethnics. The samples have been previously used for the purpose of generating reference databases at the DNA Forensic Laboratory, Department of Chemistry of Malaysia (KIMIA).

2.3.2 Forensic casework samples

Twenty-eight crime stains and thirty unknown bone extracted samples used were also from the DNA Forensic Laboratory, Department of Chemistry of Malaysia (KIMIA). Results of the samples had led to generation of either partial DNA profiles or negative results. The samples were taken from cases which had completed court trial.

2.3.3 Serial dilution samples

Two sets of serial dilution of commercial DNA from Human Genomic Female DNA (193 µg/ml) and Human Genomic Male DNA (273 µg/ml) (both controls are from Promega), respectively were prepared with different concentrations; Set 1 (10.0, 5.0, 1.25, 0.6, 0.3, 0.15, 0.07, 0.03, 0.01 and 0.009) ng and Set 2 (1.0, 0.5, 0.25, 0.125, 0.06 and 0.03) ng/µl. Set 1 was amplified using Qiagen® Investigator® DIPplex kit and in-house multiplexes. Both dilution sets were amplified using Mini-INDEL multiplex PCR. The amplification was carried out using GeneAmp® PCR System 9700 Thermal Cycler (Thermo Fisher Scientific), Applied Biosystems 2720 Thermal Cycler (Thermo Fisher Scientific) and Veriti® 96 Well Thermal Cycler (Applied Biosystems®). The PCR products were run and analyzed on Applied Biosystems® 3500 Genetic Analyzer.

2.3.4 Artificially degraded samples

Degradation of commercial Human Genomic Female DNA (193 µg/ml) (Promega,) was carried out using RQ1 RNase-Free DNase I (Promega) that has the ability to degrade both double stranded and single stranded DNA (refer to 3.4.9 and 5.3.5.3). A reaction mixture was prepared as described by (Swango et al., 2006, Asamura et al., 2007) at a total volume of 55 µl by mixing 2.1 µl of female DNA control (DNA concentration in a final volume is 7.37 ng/55 µl), 5.5 µl of RQ1 DNase 10X Reaction Buffer (400mM Tris-HCl (pH 8.0), 100 mM $MgSO_4$, 10 mM $CaCl_2$) and 47.4 µl of nuclease-free water was added to adjust the remaining volume. Before adding the 100 times diluted DNase I (0.01 U/µl) to the mixture, 5 µl was removed as a control and then 5 µl of the diluted DNase I was added to the 50 µl mixture and incubated at 24 °C. Five microliters were taken from the solution as a serial degradation at: (2, 5, 10, 20, 30, 45, 60, 90, 120 and 180) min. One microliter of RQ1 DNase Stop Solution (20 mM EGTA (pH 8.0) was added to stop the degradation reaction and incubated at 65 °C for 15 min. The degradation degree was assessed and amplified using GeneAmp® PCR System 9700 Thermal Cycler (Thermo Fisher Scientific), Applied Biosystems 2720 Thermal Cycler (Thermo Fisher Scientific) and Veriti® 96 Well Thermal Cycler (Applied Biosystems) and the PCR products were run on ABI 3500 Prism® Genetic Analyzer.

2.4 Quality control

Appropriate controls such as positive, negative and reagent blank were prepared and processed with the samples to ensure the reagents used are working well and to detect and minimize contamination. In these research, positive control used are known DNA (which supplied together in kits) and Human Genomic DNA Male/Female DNA, whilst negative control and reagent blank are just distilled water.

2.5 General DNA profiling methods

2.5.1 DNA extraction and purification

The population samples stained on FTA® card (Whatman® Bioscience) and casework samples were extracted using phenol/chloroform (organic extraction) (Sigma-Aldrich) and bone samples were extracted using PrepFiler® Forensic DNA Extraction Kit

(Thermo Fisher Scientific). The bone samples were provided by KIMIA. Reagent blanks were processed along with all extractions.

2.5.1.1 Organic extraction

The population and casework samples were extracted using the phenol/chloroform extraction method with some modification from a previous published protocol (Köchl, Niederstätter & Parson 2005) (to Section 3.3.1.2). Approximately, 10 mm^2 of FTA® card (Whatman® Bioscience) was cut into a 1.5 microcentrifuge tube. 300 µl of digestion buffer (10 mM Tris-HCl, 10 mM EDTA, 50 mM NaCl and 2% SDS, pH 7.5), 15 µl of 10 mg/ml proteinase K (Qiagen®) was added to the sample. The samples were then mixed thoroughly by vortexing and incubated at 56 °C overnight in the incubator until the sample digested completely. Following the incubation, 300 µl of phenol/chloroform/isoamyl alcohol (25:24:1, pH 8.0) (Fisher Bio Reagents) was added to the digestion buffer and mixed briefly by vortexing the mixtures to attain a milky emulsion. It was then centrifuged for 5 min at 10,000 x *g*. The aqueous (upper) layer was transferred to the pre-labelled 1.5 ml microcentrifuge tube, this step was repeated twice. Next, the aqueous layer was transferred to the Microcon® Centrifugal Filter Devices (Merck Milipore) in order to concentrate, purify and recover the extracted DNA according to the manufacturer's instructions. A final volume of approximately 40 µl was collected after the sample was centrifuged for 5 min at 7,500 x *g.* The extracted DNA was then transferred to new pre-labelled 1.5 ml microcentrifuge tube and stored at -20 °C.

2.5.1.2 Gentra® PureGene® Tissue Kit (Qiagen®).

The buccal swab used as reference samples were extracted using the Gentra® Puregene® Tissue Kit (Qiagen®) with some modification from the manufacturer's standard protocol (Gentra® Puregene® Handbook June 2011). First, the buccal swab head, which has been cut off was placed in a 1.5 ml microcentrifuge tube. Then, 300 µl of the cell lysis solution and 3 µl of proteinase K were added to the tube and it was mixed by gentle inversion. The sample was incubated at 56 °C for 30 min and the tube was gently vortexed every 10 min. After the incubation, the swab head was removed from the tube and 100 µl of protein precipitation solution was added to the tube. The mixture was then vortexed at high speed for 30 s and placed on ice for 5 min. The tube was then centrifuged for 3 min at ≥10,000 x *g* and later the supernatant was transferred into new pre-labelled 1.5 ml microcentrifuge tube. To the tube, 300 µl of isopropanol was added

and the mixture was gently mixed by inverting the tubes. The sample was incubated at -20 °C for 20 min to 30 min. Following the incubation, the tube was centrifuged for 5 min at ≥10,000 x *g* and the supernatant was discarded. Next, to the tube, 300 µl of 70% ethanol was added and the tube was gently inverted in order to wash the DNA pellet. It was then centrifuged for 1 min at ≥10,000 x *g* and the supernatant was discarded by pouring it out carefully. The tube was carefully inverted on clean absorbent paper and it was left air dried for 15 min to 30 min. Finally, 50 µl of DNA hydration was added to the tube and it was left for 15 min to 30 min in order to elute the DNA. The sample was then stored at -20 °C.

2.5.1.3 QIAamp® DNA Mini Kit (Qiagen®)

DNA extraction was carried out using QIAamp® DNA Mini Kit from buccal swabs according to manufacturer's instructions (QIAamp® DNA Mini and Blood Mini Handbook April 2010). Buccal swabs were placed in a 2 ml microcentrifuge tube and 400 µl of PBS (phosphate-buffered saline) were added. QIAGEN® proteinase K (20 mg/ml) (20 µl) and AL buffer (400 µl) were added to the samples and vortexed for 15 s. The samples were incubated at 56 °C for approximately 10 min, and centrifuged briefly to remove drops from inside the lid. 96% Ethanol (400 µl) was added to the samples, and mixed by vortexing then centrifuged to remove drops from inside the lid. The mixture (700 µl) was applied to the QIAamp® mini spin column (in a 2 ml collection tube) and centrifuged at 7656 x *g* for 1 min. The remaining mixture was applied to the same mini spin column and centrifuged at 7656 x *g* for 1 min. Buffer AW1 (500 µl) was added to the column and centrifuged at 7656 x *g* for 1 min. Buffer AW2 (500 µl) was added and centrifuged at full speed 23447 x *g*) for 3 min. The QIAamp mini spin column was placed in a new 2 ml collection tube and the collection tube with the filtrate was discarded. Centrifuged at full speed for 1 min to eliminate the chance of possible Buffer AW2 carryover. The QIAamp mini spin column was placed in a clean 1.5 ml microcentrifuge tube, the collection tube containing the filtrate was discarded. Nuclease-free water (100 µl) was added to the mini spin column and incubated at room temperature (15–25 °C) for 1 min, and then centrifuged at 7656 x *g* for 1 min. The samples were labelled and stored at -20 °C until further use.

2.5.2 DNA quantitation

There were three quantitation methods used in this thesis; Real Time PCR, Nanodrop and Qubit. No particular reason in using these methods as I see only which are available at that time.

2.5.2.1 Real-time PCR

The extracted samples were quantified using the Quantifiler® Human DNA Quantification Kit (Thermo Fisher Scientific) according to standard protocols recommended by the manufacturer (Quantifiler® Kits User Manual 2012) (refer to Chapter 3 and 5). However, the quantification was performed at a reduced total final volume of 12.5 µl with the following reaction volume: 5.25 µl of Quantifiler® Human Primer Mix and 6.25 µl and the DNA template used was 1.0 µl. Eight quantification standards (ranging from 50 ng/µl to 0.023 ng/µl) were prepared by carrying out serial dilution of 200 ng/µl Human DNA Standard in TE buffer. The samples were analysed using Applied Biosystems 7500 Real Time PCR System. The conditions are shown in Table 2-1. Internal Positive Control (IPC) results were monitored for the presence of inhibitors. Sample which were above 0.5 ng/µl, were diluted and stored at -20 °C.

Table 2-1: The real-time PCR conditions for quantification of extracted DNA samples.

Stage	Cycle	Temperature (°C)	Time
Stage 1	1	95	10 min
Stage 2	40	95	15 s
		60	1 min

2.5.2.2 NanoDrop™ Spectrophotometer 2000 (Thermo Scientific)

DNA concentrations were measured using the NanoDrop™ Spectrophotometer 2000 (Thermo Scientific) (refer to Chapter 5) that has a detection limit for ds (double strand) DNA using absorbance measurement ratio at 260 to 280 nm as an indication of sample purity, referring to the manufacturer's user manual (NanoDrop 2000/2000c Spectrophotometer V1.0 User Manual 2009). Prior to making the blank measurement, both measurement surfaces were cleaned using nuclease-free water. The Nanodrop™

software on the portable computer (PC) was opened and the Nucleic Acid application was selected, nuclease free water (1 µl) was used as a blanking buffer and loaded onto the lower measurement pedestal the sampling arm was lowered to amount the blank by selecting 'Blank' in the application. The buffer from both the top and bottom measurement surfaces was cleaned and removed using lint-free laboratory wipes (Kimberly-Clarks® Kimwipes™). Extracted DNA (1 µl) was pipetted on the lower measurement pedestal. After selecting 'Measure' in application software, DNA concentration and purity ratio were automatically calculated and the sample spectral output was reviewed to assess the quality of DNA.

2.5.2.3 Qubit® dsDNA HS (High Sensitivity) Assay Kits (Thermo Fisher Scientific)

Purified PCR products from the generation of internal amplification controls (IAC_{90} and IAC_{170}) were quantified using Qubit® dsDNA HS (High Sensitivity) Assay Kits (Thermo Fisher Scientific) (refer to Chapter 5) which is highly selective for double-stranded DNA. According to the user manual (Qubit® dsDNA HS Assay Kits Manual 2015), the Qubit® working solution was prepared by diluting the Qubit® dsDNA HS Reagent 1:200 in Qubit® dsDNA HS Buffer. 190 µl of Working Solution was added to the thin-wall 0.5 ml PCR tube for the standards (Std 1 and Std 2), then 10 µl of each standard were added into the appropriate tubes and mixed by vortexing for 3 s. Then, the purified PCR products (5 µl) were mixed with Qubit® dsDNA HS Working Solutions (195 µl) and added to the assay tubes, then vortexed for 3 s. All tubes were incubated at room temperature for about 2 min, then the concentration was measured using Qubit® 3.0 Fluorometer (Thermo Fisher Scientific™). On the home screen of the fluorometer, 'dsDNA High Sensitivity' was pressed and 'Read Standards' was selected to measure Std 1 and Std 2. After reading the Std 2, 'Run Samples' was selected, followed by the volume and unit of the samples were chosen on the assay screen. A sample tube was inserted into the sample chamber, then 'Read Tube' was selected. The results were displayed on the screen; dsDNA concentration was estimated in ng/µl.

2.5.2.4 Dilution

Based on the quantitation results, any sample that yielded a concentration of DNA that has more than 0.5 to 1.0 ng/µl was diluted with sterile deionized water down to approximately 0.5 to 1.0 ng/µl prior to amplification unless stated otherwise (refer to Chapter 3 and 5).

2.5.3 DNA Amplification

2.5.3.1 PowerPlex® 16 System (Promega)

The extracted samples were amplified using the PowerPlex® 16 System (Promega) according to standard protocols recommended by the manufacturer (PowerPlex® 16 System Technical Manual 2016) (refer to Chapter 3 and 5). However, the PCR amplification was performed at a reduced total final volume of 12.5 µl with the following reaction volume: 1.25 µl Gold Star 10x Buffer, 1.25 µl PowerPlex® 10x Primer Pair Mix, 0.4 µl Amplitaq Gold DNA Polymerase, 9.1 µl water and the DNA template used was 1.0 µl.

The amplification was carried out using Veriti® 96-Well Thermal Cycler, Applied Biosystems with the parameters shown in Table 2-2.

Table 2-2**:** PCR conditions for amplification of serial dilution control DNA (2800M) samples using PowerPlex® 16 System (Promega).

Stages & cycles	Temperature (°C)	Time
Stage 1 (1 cycle)	95	11 min
Stage 2 (1 cycle)	96	1 min
Stage 3	94 ramp 100%	30 s
10 cycles	60 ramp 100%	30 s
	70 ramp 23%	45 s
Stage 4	90 ramp 100%	30 s
22 cycles	60 ramp 100%	30 s
	70 ramp 23%	45 s
Stage 5 (1 cycle)	60	30 min
Stage 6 (Hold)	4	∞

2.5.4 Data analysis

2.5.4.1 Statistical analysis and forensic parameters

The population samples from five ethnic groups, Malays (N = 100), Chinese (N = 100), M-Indians (N = 100), Iban (N = 100) and Bidayuh (N = 100), were genotyped and analysed for the 30 INDELs plus amelogenin. The genetic structure of the populations and basic parameters of molecular diversity and including Analysis of Molecular Variance (AMOVA), allele frequency, heterozygosity, Hardy-Weinberg equilibrium (HWE), pairwise FST values and linkage disequilibrium (LD) were all assessed by Arlequin ver 3.5 software (http://cmpg.unibe.ch/software/arlequin3/) (Excoffier, Lischer 2010).

Statistical parameters to evaluate the forensic efficiency, such as match probabilities (MP), power of discrimination (PD), power of exclusion (PE) and polymorphic (PIC) were calculated using DNA Powerstats V1.2 (Promega) (Tereba 1999) for each locus and profile.

R studio software was used to perform the statistical analysis of variance (ANOVA) to study the peak heights (RFU) of the profile. Calculations of the averages (avg), standard deviation (s.d) and relative standard deviation (r.s.d %) were carried out using Excel 2013.

2.5.4.2 Snipper App Suite V2.0

The Snipper App Suite Version 2.0 (Phillips et al., 2007) is a web portal (http://mathgene.usc.es/snipper/legacylinks.php) which allowed to classify an AIM (Ancestry Informetaive Markers) profile as fitting to one of several populations (e.g. Europe, East Asia, Africa, America and Oceania). The portal also able to carry out some complementary tasks like plotting certain populations and profile, design an optimal training set, and simulating profiles from the training set file.

2.6 Validation of INDEL

2.6.1 Optimisation of Investigator® DIPplex Kit (Qiagen®)

Amplification process was performed using the Investigator® DIPplex Kit (Qiagen®) according to standard protocols recommended by the manufacturer (Investigator DIPplex Handbook 2010) but at a reduced total final reaction volume of 5 µl with the following reaction volume: 1.0 µl Reaction Mix A, 1.0 µl Primer Mix DIPplex, 0.12 µl MultiTaq2 DNA Polymerase and 1.88 µl of Nuclease-free water. The DNA template used was 4.0 µl and each amplification reaction tube contained 1.0 ng of DNA template. Investigator® DIPplex Kit amplifies DNA at 30 loci plus amelogenin with four different labels and they are as follows:

- 6-FAM™: amelogenin, HLD77, HLD45, HLD131, HLD70, HLD6, HLD111, HLD58, HLD56
- BTG: HLD118, HLD92, HLD93, HLD99, HLD88, HLD101, HLD67
- BTY: HLD83, HLD114, HLD48, HLD124, HLD122, HLD125, HLD64, HLD81
- BTR: HLD136, HLD133, HLD97, HLD40, HLD128, HLD39, HLD84

The PCR was performed on the Applied Biosystems GeneAmp® PCR System 9700.

2.6.2 Determination of minimum reaction volume

Table 2-3 shows several PCR reaction mix volumes used in this study to observe the sensitivity of this assay.

Component	PCR Reaction Mix Volume (µl/well)				
	Original	One half	One third	One fourth	One fifth
Reaction Mix A*	5.0 µl	2.5 µl	1.6 µl	1.25 µl	1.0 µl
Primer Mix DIPplex	5.0 µl	2.5 µl	1.6 µl	1.25 µl	1.0 µl
Multi Taq2 DNA Polymerase	0.6 µl	0.3 µl	0.2 µl	0.15 µl	0.12 µl
Nuclease free water	9.4 µl	4.7 µl	3.3 µl	2.35 µl	1.88 µl
DNA template (0.2 to 0.5 ng)	5.0 µl	2.5 µl	1.6 µl	1.25 µl	1.0 µl
Total	25.0 µl	10 µl	6.7 µl	5.0 µl	4.0 µl

* Contains dNTP mix, MgCl2, and bovine serum albumin (BSA).

Standard cycling protocol, recommended for all DNA samples as shown in Table 2-4.

Table 2-4 shows standard PCR cycles which recommended by manufacturer (adapted from Investigator® DIPplex Kit handbook):

PCR stages		Temperature (°C)	Time (min)
Initial incubation		94	4
Denaturation	30*	94	0.5
Annealing		61	2
Extension		72	1.25
Final extension		68	60
Hold		10	∞

*If stains containing small amounts (<100pg) of DNA, recommended to increase the PCR cycles to 32.

2.6.3 Fragment Analysis by capillary electrophoresis

Typical loading mix comprised of Hi-Di™ Formamide and GeneScan™ -500 LIZ Size Standard (Thermo Fisher Scientific) prepared based on the following reaction volume: 9.5 µl of Hi-Di™ Formamide and 0.5 µl of GeneScan™ -500 LIZ Size Standard. The internal size standard provided 16 single stranded labelled fragments ranging from 35 bp to 500 bp. To each well, 1.0 µl of amplified PCR product was added to 10.0 µl of loading mixture for automated analysis on the Applied Biosystems 3500 Genetic Analyzer (Applied Biosystems), run module (50_POP6) was selected with dye sets BT5: 6-FAM (blue), BTG (green), BTY (yellow), BTR (red) and BTO (orange). The parameters and instrument operating procedures are followed with that given in the Investigator® DIPplex Handbook (Qiagen®) (Investigator DIPplex Handbook 2010) which is shown in Table 2-5.

Table 2-5: The parameters of ABI 3500_POP6 module.

Parameter	Setting
Oven temperature (°C)	60
Run voltage (kV)	15
Pre-run voltage (kV)	15
Injection voltage (kV)	3.0
Injection time (s)	10.0
Run time (s)	2700

In order to generate a DNA profile, the amplified PCR fragments were analysed with GeneMapper™ IDX software V1.2 (Applied Biosystems™). The parameters for the analysis DNA profiles were kept the same for every run (refer Table 2-6)

Table 2-6: Shows the parameter for the analysis of PCR fragments.

Parameters	Settings
Peak detection algorithm	Advanced
Ranges	Analysis: partial range Start point: 1000; stop point 20,000 Sizing: All sizes
Smoothing and baselining	Smoothing: light Baseline window: 51 pts
Size calling method	Local Southern Method
Peak detection	Peak amplitude thresholds 50 RFU Polynomial degree: 3 Peak window size: 11 pts Slope thresholds: 0.0

.

2.7 Development of Mini 4-plex & Internal Amplification Controls

This project was carried out by two members of the Forensic Genetics Group, UCLAN that includes MRes student, Balnd Mustafa Albarzinji, from April 2014 to September 2015 and myself, jointly undertook a project. The Mini 4-plex assay is to utilize fragments at 50 bp, 70 bp, 112 bp and 154 bp to assess degradation, whilst 90 bp and 170 bp DNA template are used as Internal Amplification Controls (IACs) for the assessment of PCR inhibition.

2.7.1 Primer design

2.7.1.1 Mini 4-plex amplicons

Sequence data for the recombination activating genes (RAG) 1 and 2 from human and pig were downloaded from Genbank (http://www.ncbi.nlm.nih.gov/genbank/) which as Table 2-7.

Table 2-7 shows the sequence data for RAG 1 and 2 for both human and pig.

Origin	RAG	Website address
Human	RAG-1	http://www.ncbi.nlm.nih.gov/gene/5896
	RAG-2	http://www.ncbi.nlm.nih.gov/gene/5896
Pig	RAG-1	http://www.ncbi.nlm.nih.gov/nuccore/AB091392.1
	RAG-2	http://www.ncbi.nlm.nih.gov/nuccore/AB091391.1

The sequences for RAG-2 from human and pig were aligned to identify conserved regions for designing primers of 50 bp and 70 bp amplicons. The sequences for RAG-1 from human and pig were aligned to identify conserved regions for designing primers of 112 bp and 154 bp amplicons by using BioEdit Software V7.2. Identical regions between human and pig were selected and primers for amplicons 50 bp, 70 bp, 112 bp, 154 bp from RAG-1 and RAG-2 were designed using the publicly available software Primer3 (http://primer3.ut.ee/) (refer to Table 2-8). The self-complementarity of individual primers was assessed using the NCBI (National Centre for Biotechnology Information) (http://www.ncbi.nlm.nih.gov/) and Primer-BLAST (Basic Local Alignment Search Tool) programs (www.ncbi.nlm.nih.gov/Entrez/). The annealing temperatures for the four pairs of the primers were 60 °C ± 2 and the GC% was in the standard range of 40-60 %. Primers were synthesized by Invitrogen™ (Thermo Fisher Scientific), the forward primers

were labelled with fluorescein dye at 5' end, while, the reverse primers were unlabelled. All primer pairs were synthesized, purified using HPLC, lyophilized and delivered. Stock solutions (100 µM) were prepared by adding the appropriate volume of the nuclease-free water and stored at – 20 °C. Aliquots of working solution (10 µM) were prepared (10 µl from 100 µM to 90 µl of nuclease-free water) and stored at 4 °C for regular use.

Table 2-8: Primer sets targeting conserved regions of human and pig (refer to Chapter 5).

Nuclear gene	Amplicon length (bp)	PCR primers forward and reverse (5'-3')	Melting Temperature (°C)
RAG-2	50	F: TGGATTACATGCTGCCCTACT R: TGGTACCCAAGTGTTGATATCCA	58 58
RAG-2	70	F: ACCCAGCCACTTGCACATT R: TTTCCCTCCATGGATGATGT	60 59
RAG-1	112	F: GAGGGAGCTCAAGCTGCAA R: GTGCTCATTCCTCGCCCT	60 59
RAG-1	154	F: TCGGGGACTCAAGAGGAAGA R: GCAGTTGGCGATCTTCTTCA	59 58

2.7.1.2 Internal amplification control (IAC_{170})

Two internal amplification controls had previously been designed from different regions of the plasmid PBR322 to amplify IAC_{90} and IAC_{410} using tailed primers in the first PCR that are the binding sites of the second sets of the primers in the second PCR (Fondevila et al. 2012). New reverse primer was designed to amplify an amplicon 170 bp with ROX-labelled forward primer of IAC_{410} in the second PCR after generation of the IACs in the first PCR. Nuclear sequence data for the DNA Plasmid pBR322 was downloaded from Genbank (http://www.ncbi.nlm.nih.gov/nuccore/208958) (refer to Table 2-9). The reverse primer was designed using the publicly available software Primer3 (http://primer3.ut.ee/); the melting temperature was 59 °C and the GC % was 52.38 % in the standard range 40-60 %. The primer was synthesized, desalted and delivered by Invitrogen™ (Thermo Fisher Scientific). Stock solution (100 µM) was prepared by adding the appropriate volume of the nuclease-free water and stored at – 20 °C. A working solution aliquot (10 µM) was prepared and stored at 4 °C for regular use.

Table 2-9: Primer sequences of the generation of IAC_{90} and IAC_{410} in the first PCR and ROX labelled primers used in the second PCR.

	1st PCR [a] (Sequence 5' to 3')
IAC_{90}	F: **CTGTCAAATCTAAACACCCTGATGCG**GCTTGCGGTATTCGGAATCTTG
	R: **GTCAGCTTGCATAATATCGAGATAACGC**GAGCGAGGGCGTGCAAGATT
IAC_{410}	F: **TGTCAAATCTAAACACCCTGATG**CGGATGCTGCTGGCTACCCTGT
	R: **GTACAATGTTGACGTTCCTCGCTG**CGTGAAGCGATTCACAGATCTCTG
	2nd PCR [b.] (Sequence 5' to 3')
IAC_{90}	F: **CTGTCAAATCTAAACACCCTGATGCG**
	R: GTCAGCTTGCATAATATCGAGATAACGC
IAC_{410}	F: **CTGTCAAATCTAAACACCCTGATGCG**
	R: GTACAATGTTGACGTTCCTCGCTG
IAC_{170}	F: **CTGTCAAATCTAAACACCCTGATGCG**
	R: GATGAACATGCCCGGTTACTG

Note: (a) The bold sequences are the tailed primers in the first PCR. (b) The red coloured letters showed the ROX labelled primers in the second PCR and the sequences are the same for IAC_{90}, IAC_{70} and IAC_{410}.

2.7.2 Singleplex PCR optimisation of the new markers

2.7.2.1 ThermoPrime 2x ReddMix PCR Master Mix (1.5mM $MgCl_2$) (Thermo Fisher Scientific)

Primer sets were tested to amplify the fragments of the Mini 4-plex individually using a ready-to-use ThermoPrime 2X ReddyMix PCR Master Mix (1.5 mM $MgCl_2$) (Thermo Fisher Scientific). 1 µl of extracted DNA from a buccal swab (4 ng/µl) was used as a control template in a PCR for a reduced final reaction volume of 15 µl containing: 0.5 µl of each forward and reverse of appropriate concentration (0.5 µM), 6 µl of 2X Reddy MiX™ PCR master mix containing ThermoPrime *Taq* DNA Polymerase (0.625 units), Tris-HCl- (pH 8.8 at 25 °C (75 mM), $(NH_4)_2$ SO_4 (20 mM), Tween® 20 (0.01 % v/v), each dNTP (0.2 mM) and 1.5 mM $MgCl_2$. 7 µl of nuclease-free water was added to adjust the remaining volume. Gradient PCR (different annealing temperatures 56 °C, 58 °C, 60 °C

and 62 °C) as shown in Table 2-10 was used to test the primers using Veriti® 96-Well Thermal Cycler Applied Biosystems.

Table 2-10: PCR conditions for amplification of the new markers individually using Thermo Prime 2x ReddyMix PCR Master Mix (1.5mM $MgCl_2$) (Thermo Fisher Scientific).

PCR stages .		Temperature (^oC)	Time
Initial incubation		95	2 min
Denaturation	35	95	25 s
Annealing		56-62*	35 s
Extension		72	1 min
Final extension		72	5 min
Hold		4	∞

*Annealing temperatures at 56 °C, 58 °C, 60 °C and 62 °C were tested with the new primers.

2.7.2.2 Platinum® PCR SuperMix High Fidelity (Thermo Fisher Scientific)

Primer sets for the markers of the mini 4-plex were tested individually using Platinum® PCR SuperMix High Fidelity (Thermo Fisher Scientific) to compare the results with ThermoPrime 2X ReddyMix PCR Master Mix (Thermo Fisher Scientific). The component of the Platinum® PCR SuperMix is 22 U/ml complexed recombinant *Taq* DNA polymerase, *Pyrococcus* species GB-D thermostable polymerase, and Platinum® *Taq* Antibody; 66 mM Tris-SO_4 (pH 8.9); 19.8 mM $(NH_4)_2SO_4$; 2.4 mM $MgSO_4$; 220 µM dNTPs; and stabilizers. A Mixture of the proofreading enzyme, *Pyrococcus* species GB-D polymerase, with *Taq* DNA polymerase increases fidelity. Final PCR volume 15 µl was used which containing: Platinum® PCR SuperMix SuperMix (13 µl), 0.5 µl of appropriate concentration (0.5 µM) of each forward and reverse primer and DNA template 1 µl. Gradient PCR with annealing temperatures of 56 °C, 58 °C, 60 °C and 62 °C were used to test the primers using Veriti® 96-Well Thermal Cycler (Applied Biosystems).

Table 2-11: PCR conditions for amplification of the mini 4-plex individually using Platinum® PCR SuperMix High Fidelity (Thermo Fisher Scientific) (refer to Chapter 5).

PCR stages		Temperature (°C)	Time
Initial incubation		94	2 min
Denaturation	35	94	30 s
Annealing		56-62*	30 s
Extension		68	1 min
Hold		4	∞

*Annealing temperatures at 56 °C, 58 °C, 60 °C and 62 °C were tested separately with the new primers.

2.7.3 Multiplex PCR Amplification of the Mini 4-plex

Amplification of the Mini 4-plex markers as a multiplex was carried out using Platinum® Multiplex PCR Master Mix (Thermo Fisher Scientific). Respectively, (1 µl) of Human Genomic Female DNA (Promega) and DNA extracted from pig tissue was used as a control template in the PCR for a reduced final reaction volume of 10 µl comprising; PCR master mix 5 µl, primer mixture 0.6 µl, which prepared as a separate reaction using different concentrations of the primers to balance the height of the peaks, and nuclease-free water 3.4 µl was added to adjust the remaining volume. The multiplex PCR amplification was performed using the following parameters based on analysis by capillary electrophoresis as shown in Table 2-12.

Table 2-12: PCR conditions for amplification of the mini 4-plex individually using Platinum® Multiplex PCR Master Mix (Thermo Fisher Scientific).

PCR stages		Temperature (°C)	Time
Initial incubation		95	2 min
Denaturation	28	95	30 s
Annealing		60	90 s
Extension		72	30 s
Final extension		60	30 min
Hold		4	∞

The parameters were used for analyzing the results using gel electrophoresis were shown in Table 2-13.

Table 2-13: PCR conditions for amplification of the mini 4-plex individually using Platinum® Multiplex PCR Master Mix (Thermo Fisher Scientific).

PCR stages		Temperature (°C)	Time
Initial incubation		95	2 min
Denaturation	**35**	95	30 s
Annealing		60	30 s
Extension		72	15 s
Final extension		60	10 min
Hold		4	∞

Both amplifications were carried out using GeneAmp® PCR System 9700 Thermal Cycler.

Table 2-14: The concentration and volume of the oligonucleotides as a mixture (total volume 84 µl) to amplify the mini 4-plex multiplex PCR, prepared from 10 µM working solutions (refer to Chapter 5).

PCR Primers	Primers volume in a mixture (µl) (10 µM)	Primer concentrations in a mixture (µM)	Primer concentrations in a PCR final volume of 10 µl (µM)
50 bp_F	10	1.2	0.072
50 bp_R	10	1.2	0.072
70 bp_F	11	1.3	0.078
70 bp_R	11	1.3	0.078
112 bp_F	10	1.2	0.072
112 bp_R	10	1.2	0.072
154 bp_F	11	1.3	0.078
154 bp_R	11	1.3	0.078

2.7.3.1 Species specificity

Mini 4-plex PCR amplifications was performed on pig, sheep, roe, rabbit and rat samples. Pig DNA was extracted from of meat samples using DNeasy® Blood and Tissue Kit (Qiagen®) (refer to 2.7.3.2). Extracted DNA for sheep, roe, rabbit and rat was provided by Dr. Arati Iyengar (Lecturer in UCLAN). All samples were diluted to 1 ng/ µl.

2.7.3.2 Pig tissues

Extraction of DNA from pig tissue samples was performed using DNeasy® Blood and Tissue Kit (Qiagen®). Pig tissue (25 mg) was cut into small pieces, and placed in a 1.5 ml microcentrifuge tube. Then 180 µl of ATL buffer was added. 20 µl of Proteinase K (20 mg/ml) was added and mixed by vortexing then incubated at 56 °C until the tissue was completely lysed. AL buffer (200 µl) and 200 µl of ethanol (96 %) were added to the samples and mixed by vortexing. The samples were transferred into the DNeasy mini spin column and centrifuged at 7656 x *g* for 1 min. 500 µl AW2 buffer was added and centrifuged for 3 min at 23,447 x *g*. The DNeasy, mini spin column was placed in a clean 1.5 ml microcentrifuge and 150 µl of AE buffer was pipetted onto the membrane, incubated at room temperature for 1 min and centrifuge for 1 min at 7656 x *g.* The samples were labelled and stored at -20 °C.

2.7.4 Development of Internal Amplification Control (IACs)

2.7.4.1 Generation of IACs in the first PCR

Generation of the two markers from pBR322 Plasmid DNA (0.5 µg/ µl) (Thermo Fisher Scientific™) was carried out using ThermoPrime 2X ReddyMix PCR Master Mix (1.5 mM $MgCl_2$) (Thermo Fisher Scientific) in a reduced final volume of 15 µl comprising: 6 µl of 2X Reddy MiX™ PCR master mix, 7 µl of nuclease-free water, 1 µl of diluted plasmid (12 ng/µl) and 0.5 µl of 10 µM of each forward and reverse primers. The amplification was carried out using Veriti® 96-Well Thermal Cycler (Thermo Fisher Scientific) and the optimized conditions shown in Table 2-15 were used. The samples were hold at 4 °C until removed from the PCR tool.

Table 2-15: PCR conditions for generation of IAC90 and IAC410 from the plasmid pBR322 using Thermo Scientific 2X ReddyMix PCR Master Mix (1.5 mM $MgCl_2$) (refer to Chapter 5).

PCR stages		Temperature (°C)	Time
Initial incubation		95	2 min
Denaturation	35	95	30 s
Annealing		66 (IAC_{90}) 58 (IAC_{410})	30 s
Extension		72	1 min
Final extension		72	5 min
Hold		4	∞

2.7.4.2 Purification of IAC_{90} and IAC_{410} PCR products

Purification of the PCR products was carried out using QIAquick® PCR Purification Kit (Qiagen®). Buffer PB was added to the PCR products (75 µl) (5 volumes of the PCR reaction). The mixture was applied to the QIAquick® column and centrifuged at 13,000 RPM for 1 min. Buffer PE (750 µl) was added to wash the products and centrifuged at 20217 x *g* RPM for 1 min. The QIAquick® column was placed in a clean 1.5 ml microcentrifuge tube and 50 µl of Buffer EB (10 mM Tris·Cl, pH 8.5) was added to elute the DNA. The purified PCR products were stored at -20 °C until further use.

2.7.4.3 Amplification of the IAC_{90} and IAC_{170} in the second PCR

The IACs purified PCR products were diluted 10-fold to optimise the concentration for balanced peaks on electropherograms. The amplification was performed using Platinum® Multiplex PCR Master Mix (Thermo Fisher Scientific). In a reduced total volume of 10 µl comprising: Platinum® master mix. (5 µl), two pairs of ROX-labelled forward primer and unlabelled reverse primers as a mixture (0.6 µl), IACs template (1 µl) and nuclease-free water (3.4 µl). The amplification was shown in Table 2-16 and carried out using Veriti® 96-Well Thermal Cycler Applied Biosystems, the PCR products were run and analysed on ABI 3500 Genetic Analyser.

Table 2-16: PCR conditions of IAC_{90} and IAC_{170} amplification as a multiplex in the second PCR.

PCR stages		Temperature (°C)	Time
Initial incubation		95	2 min
Denaturation	35	95	30 s
Annealing		60	90 s
Extension		72	30 s
Final extansion		60	30 min
Hold		4	∞

2.7.5 Electrophoresis

2.7.5.1 Gel electrophoresis

The PCR products were analysed by agarose gel electrophoresis (AGE) for both singleplex and multiplex PCR. AGE (2 % (W/V)) was prepared in an Erlenmeyer flask (Sigma Aldrich) using 1 g of Molecular Biology (MB) Grade Agarose Powder (Severn Biotech Limited) added to 50 ml of 1X TAE buffer (per 100 ml: 4.84 g Tris Base, 2 ml 0.5 EDTA (pH 8.0) and 1.142 ml glacial acetic acid). 2 µl of Safe View Nucleic Acid Stain (NBS Biologicals Limited) or GelRed™ Nucleic Acid Gel Stain (Biotium) were used for DNA visualization under ultraviolet light using the BioDoc-It® Imaging System Benchtop 2UV Transilluminator. The warm gel was poured in the mini agarose gel tray (EmbiTech UVT), the gel then was submerged into 1X TAE buffer in the electrophoresis tank. The PCR products were prepared as follows: 4 µl of the product was added and mixed

thoroughly with 2 µl of 6X loading buffer and loaded into the well. 3 µl of 50 bp or 100 bp DNA ladders (NBS Biologicals Ltd, UK) were used and loaded into the well as a size standard. The gel was run on 100 V for 20 min.

2.7.5.2 Capillary electrophoresis

Capillary electrophoresis (CE) ABI 3500 Prism® Genetic Analyzer (Applied Biosystems) was used to perform the DNA fragment analysis of the mini 4-plex multiplex PCR products. The fragment analysis 50_POP6 (Array separation distance 50 cm and polymer type POP-6) was used with the dye set DS-33 (filter set G5): 6-FAM™ (blue), VIC® (green), NED™ (yellow), PET® (Red) and LIZ® (orange). The analysis was carried out using the following parameters: run temperature 60 °C, pre-run voltage 15 KV, pre-run time 180 s, injection time 5 s, injection voltage 1.6 kV, run voltage 15 kV and run time 2700 s. The samples were prepared by mixing (8.5 µl) Hi-Di™ Formamide (Thermo Fisher Scientific), 1 µl PCR products and 0.5 µl GeneScan™ –500 LIZ® Size Standard (Thermo Fisher Scientific) and placed into appropriate wells of the MicroAmp Optical 96-well Reaction Plate (Thermo Fisher Scientific). The prepared samples were denatured at 94 °C for 4 min and cooled for about 7 min before start running using capillary electrophoresis. The data were analyzed by using GeneMapper® ID-X Software V1.2 (Applied Biosystems™), with a constant parameters shown in Table 2-17.

Table 2-17: The parameters of GeneMapper® ID-X Software (version 1.2) for data analysis.

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

2.8 Development of Mini INDEL assay

2.8.1 Primer design

Table 2-18: The redesigning of Mini-INDEL primers from (Pereira et al. 2009) showing the length, TM, GC and sequences from 5′ to 3′

Marker		Length	TM	GC%	5′ to 3′
B3	F	21	58.94	52.4	AAACCTCCTCCCTCAGAAGAG
	R	20	60.93	50.0	ATTGGGTATCCCCTCGTTTG
B4	F	21	58.76	47.6	ACTGC GTTTC TGTAG AGGAG T
	R	28	58.52	28.6	AACATCCAAATTAGCTTCACATTTTTTC
B5	F	19	59.81	57.9	TCAGTGAACCTGGGACAGC
	R	24	57.16	54.3	GGAATCCCCATTCTTTTACTGAC
G6	F	25	59.80	40.0	CACTAAACTCTTCTTCTTGCAGCTT
	R	22	57.58	45.5	GGAACAGGACCATAGCATAACT
G8	F	27	58.85	37.0	GGACTTTAGTAGAAGAGGAAAATACCA
	R	24	59.56	37.5	AAATGAGACCCTTCTTAGGTTCAA
Y4	F	19	58.38	57.0	GCCTGAATAGATGCACCCG
	R	24	58.06	37.5	AATGTTCTAGAATCCACAAAGAGC
Y5	F	27	57.42	29.6	ACCATTAATAATAAAGTGTGGAAAGAC
	R	31	56.61	25.8	TTTATAAATACATAAAACGTATGAGCTAAC
R3	F	25	56.48	36.0	CATAAAGCAACTCTATTCCTTTTCC
	R	28	60.64	39.2	AATAGTCCAGTCTACCCAAATGTATTCC
R5	F	25	57.64	36.0	CTGATTCTGTTCTTTGTAATTCGGA
	R	27	59.27	32.0	TTATATACAAAGCAAAGGATGCTCA

2.8.2 Primer synthesis and purity

Each primers pairs with 5' fluorescein labelled forward primers and unlabelled reverse primers were synthesized by Thermo Fisher Scientific and were delivered in HPLC purified, desalted and lyophilised. Stock solutions (100 µM) were prepared by adding the appropriate volume of the nuclease-free water and stored at – 20 °C. Aliquots of working solution (10 µM) were prepared (10 µl from 100 µM to 90 µl of nuclease-free water) and stored at 4 °C for regular use.

2.8.3 PCR mixture

Table 2-19: The PCR reaction mix volumes used in this study to observe the sensitivity of this assay (refer to Chapter 7).

Component	PCR Reaction Mix Volume (µl/well)				
	MM1	MM2	MM3	MM4	MM5
Platinum® PCR SuperMix High Fidelity	12.5	12.5	5.0	12.5	10.0
Primer Mix	5.0	1.5	0.5	2.5	1.5
DNA template (0.5 ng/µl)	1.0	3.0	1.0	10.0	1.0
dH_2O	6.5	8.0	3.5	-	-
Final volume of master mix	25.0	25.0	10.0	25.0	12.5

2.8.4 PCR amplification

Primers were synthesized by Applied Biosystems and all forward primers were labelled with either FAM™, NED™, VIC™ and PET™ dyes which enabled colour separation on ABI 3500 Genetic Analyzer. All reverse primers were unlabelled and had additional 5' guanine base added to produce adenylated PCR products (Hill et al. 2008). The primer sequences and final concentrations of each locus used in this study presented in Table 2-20.

Amplification of samples were performed in reaction volumes of 12.5 µl; using a master mix containing 10.0 µl Platinum Master Mix (Thermo Fisher Scientific), 1.0 µl of each forward and reverse primer with 0.5 µM concentration respectively, 1.5 µl of DNA

template. Amplification reactions were performed using an ABI 2720 Thermal Cycler Life Technologies. Thermal cycling condition for Mini-INDELs was; 95 °C for 2 min; 30 cycles of 95 °C for 30 s, 58 °C for 1 min and 72 °C for 1 min followed by final incubation at 60 °C for 60 minutes. These parameters are also the same apply for New Mini-INDELs multiplex (combination of the new Mini-INDELs with the previous Mini-INDELs).

Table 2-20: The concentration of the oligonucleotides as a primer mixture (total volume 18 µl) to amplify the mini-INDELplex PCR, prepared from 10 µM working solutions (refer to Chapter 7).

PCR Primers	Primer concentrations in a primer mix (µM)	Primer concentrations in a PCR final volume of 12.5 µl (µM)
B3	1.1	0.13
B4	0.5	0.06
B5	0.5	0.06
G8	0.5	0.06
G6	0.5	0.06
Y4	0.5	0.06
Y5	1.1	0.13
R5	1.1	0.13
R3	1.1	0.13

2.8.5 Capillary electrophoresis

The parameters of the capillary electrophoresis (CE) ABI 3500 Prism® Genetic Analyzer (Applied Biosystems) used on the Mini-INDELs multiplex PCR products is the same as described at paragraph 2.7.5.2.

CHAPTER 3

EVALUATION OF INSERTION/DELETION POLYMORPHISMS (INDELS) IN THE MALAYSIAN POPULATION AND FORENSIC CASEWORK SAMPLES

3.1 Overview

This chapter will focus on the evaluation of sensitivity, reproducibility and robustness of the Qiagen Investigator DIPplex® Kit. These can be useful as a supplementary method to STR typing and SNPs. INDEL markers are analysed using short amplicons, under 160 bp, and are therefore more likely to be successful when analysing degraded samples. Furthermore, this technique is very easy, straight forward and fast to analyse using the same methods and technologies as STR.

The internal validation and optimisation of the Qiagen Investigator DIPplex® Kit was performed according to the standards of the Scientific Working Group (SWGDM) (Scientific Working Group on DNA Analysis Methods (SWGDAM),2004) to ensure maximum detection and amplification of the DNA. Some common validation criteria included: (1) establishing that the DNA typing system is sensitive and performs consistently using DNA control, (2) that identical results are obtained irrespective of the type of tissue from which DNA was extracted, (3) that the system yield consistent results, and (4) that the system performs well when used to analyse samples similar to those encountered in forensic casework (Leat et al., 2004).

3.2 Aims and objectives

- To evaluate by investigating the sensitivity, reproducibility and robustness of the kit.
- The 30 loci in the Qiagen Investigator DIPplex® kit were evaluated in five Malaysian populations i.e. the Malays, Chinese, M-Indian, Iban and Bidayuh.
- To optimise and validate the INDEL system for processing of crime scene, bone and reference samples.

3.3 Material and methods

3.3.1 Purification of DNA on FTA paper

To investigate the effect of different DNA purification methods on the quality of the DNA results, DNA was purified using two different methods prior amplification; wash with alkaline and phenol chloroform. In order to save samples, only five bloodstained specimens from the Malay population were used for each purification methods.

3.3.1.1 The purification of DNA from blood dried on FTA paper using an alkaline wash procedure

Five Malay blood-stained on FTA cards (Whatman) were purified using a method developed by the Forensic DNA Laboratory, Department of Chemistry Malaysia, followed by amplification that was described in Section 2.6.1. A 1.2 mm punch of FTA card was placed into tube and washed with 10 mM NaOH followed by TE buffer, to remove haemoglobin and other inhibitors of the PCR reaction. Once the red colour haemoglobin was removed, the punched FTA was added directly to the PCR reaction. The PCR master mix used was 5 µl.

3.3.1.2 Phenol chloroform

Another five blood-stained on FTA cards (Whatman) were extracted using the phenol chloroform method (as described in section 2.5.1.1, Chapter 2) and the supernatant was quantified and used for amplification. The total volume for PCR volume used for each samples was 5 µl.

3.3.2 Forensic casework samples

Twenty-eight crime stains and thirty unknown bone extracted samples used were also from the DNA Forensic Laboratory, Department of Chemistry of Malaysia (KIMIA). Results of the samples had led to generation of either partial DNA profiles or negative results. The samples were taken from cases which had completed court trial.

3.3.3 PCR amplification

Firstly, PCR amplification was performed as per manufacturer's recommendations. Amplification was performed in a final volume of 25 µl containing of 5 µl 0.3 to 0.5 ng

DNA template, 5 µl Reaction Mix A, 5 µl Primer Mix DIPplex®, 0.6 µl Multi Taq2 DNA polymerase and 9.4 distilled water (Investigator DIPplex Handbook 2010). PCR amplification was then performed in a final volume that was half, one third, one fourth and one fifth of the manufacturer's recommendations. Amplification took place in final volumes of (10.0, 6.7, 5.0 and 4.0) µl (refer to Table 3-1). All samples were prepared in duplicate and amplified on two different thermal cyclers; Applied Biosystems Veriti™ 96-Well Thermal Cycler and Applied Biosystems 2720 Thermal Cycler (Life Technologies).

Table 3-1 shows several PCR reaction mix volumes used in this study to observe the sensitivity of this assay.

Component	PCR Reaction Mix Volume (µl/well)				
	Original	One half	One third	One fourth	One fifth
Reaction Mix A*	5.0 µl	2.5 µl	1.6 µl	1.25 µl	1.0 µl
Primer Mix DIPplex®	5.0 µl	2.5 µl	1.6 µl	1.25 µl	1.0 µl
Multi Taq2 DNA Polymerase	0.6 µl	0.3 µl	0.2 µl	0.15 µl	0.12 µl
Nuclease free water	9.4 µl	4.7 µl	3.3 µl	2.35 µl	1.88 µl
DNA template (0.2 to 0.5 ng)	5.0 µl	2.5 µl	1.6 µl	1.25 µl	1.0 µl
Total	25.0 µl	10 µl	6.7 µl	5.0 µl	4.0 µl

3.3.4 Allelic ladder

An allelic ladder is a DNA sample consisting of all possible alleles. It is used for correct sizing of DNA fragments. A volume of 0.5 and 1.0 µl allelic ladder DIPplex® kit were loaded separately with batch of samples prepared for capillary electrophoresis.

3.3.5 Size standards

A cocktail consisting of size standard and HiDi formamide was prepared by adding 0.5 µl of GeneScan™-500 LIZ® and 8.5 µl of HiDi formamide per sample.

3.3.6 Sensitivity study

To investigate the robustness and reproducibility of the system, commercial DNA Human Genomic Female Control (Promega) was analysed in triplicate at total amounts between 10 ng to 0.009 ng of DNA per PCR. The serial dilutions were carried out as described in Section 2.3.3. Genotypes of the diluted control DNA samples for all markers were confirmed in a single reaction. Correctly typed genotypes as well as missing alleles and allelic drop-ins for each sample and DNA quantity were recorded and compared over all analysis.

3.3.7 PCR cycles for low template DNA

The manufacturer recommended to increase PCR cycles to 32, for DNA template <100 pg. The diluted Human Genomic Female DNA (Promega) which produced partial profiles (I = 0.03 ng, J = 0.01 ng and K = 0.009 ng) with 30 cycles were amplified using 32 cycles (refer Table 3-2). However, the PCR master mix volume remained the same, 5 µl.

Table 3-2 shows 32 PCR cycles which recommended by manufacturer (adapted from (Investigator DIPplex Handbook 2010) for stains containing small amounts (<100pg) of DNA.

PCR stages		Temperature (°C)	Time (min)
Initial incubation		94	4
Denaturation	32	94	0.5
Annealing		61	2
Extension		72	1.25
Final extension		68	60
Hold		10	∞

3.3.8 DNA degradation study

The performance of the assay in cases of low quality DNA was assessed by analysing artificially degraded DNA samples comparing the observed INDEL profiles with the profiles obtained using a routine forensic STR kit. Commercially available control, Human Genomic Female Control (Promega), was degraded using an adopted protocol for digestion with Promega RQ1 RNase-Free DNase I described by (Swango et al., 2006,

Asamura et al., 2007) (see section 2.3.4, Chapter 2). The degraded DNA was analysed with Qiagen Investigator DIPplex® Kit and the PowerPlex® 16 system (Promega) according to modified protocol.

3.3.9 Quality of DNA profile

In this study, assay performance was tested on Human Genomic Male Control (Promega) at 1 ng to 0.015 ng of DNA template. The number of INDEL profiles that exhibited full, partial, or no profiles, intra-colour balance (ICB), peak height ratio (PHR) and allele designation were assessed. The analysis was performed using Genemapper® ID-X Software V1.2 (Applied Biosystems) and the minimum peak height threshold was set at 50 RFU for all dyes. The purpose of setting the threshold at 50 RFU was to allow for detection of all peaks clearly above background levels.

The calculations for ICB and PHR were based on those described in (Wang et al., 2011). The ICB was calculated by first averaging heterozygous peak heights and dividing homozygous peak heights by 2. Once normalized for diploid, the lowest score for a locus labelled with a given dye was divided by the highest and the result reported as a percentage It provides an indication of the marker-to-marker colour balance within a dye channel. However, the peak height ratio (PHR), provides a measure of the balance between alleles within a locus. The PHR is calculated as the ratio of the minimum peak height to maximum peak height between two heterozygous alleles and the result reported as a percentage.

3.4 Results

3.4.1 Reaction volume of multiplex

In order to test the sensitivity of the Qiagen Investigator DIPplex® kit, dilution series of the Positive Control 9948 (500 pg/µl) were typed in 5 µl (one fifth) of PCR reaction volume described in Table 3-1. The amount of DNA used ranged from 62 pg to 2000 pg (2 ng) and each on duplicate runs. According to (Larue et al. 2012), 25 µl of PCR reaction volume gave reproducible results for all series of DNA concentrations, but in this study, the final composition of 5 µl reaction mix was chosen due to its sensitivity as well as a reduced amount of reaction mix used in this system.

The Positive Control 9948 were successfully amplified using four amplification protocols mentioned above. No differences in the results from the two different thermal cyclers observed; Applied Biosystems Veriti™ 96-Well Thermal Cycler and Applied Biosystems 2720 Thermal Cycler (Life Technologies). The peak heights obtained from the DNA profiles generated (see Figure 3-1) were above level of detection (LOD) 50 relative fluorescent unit (RFU), therefore despite a reduction in reaction volume adequate profiles were still obtained.

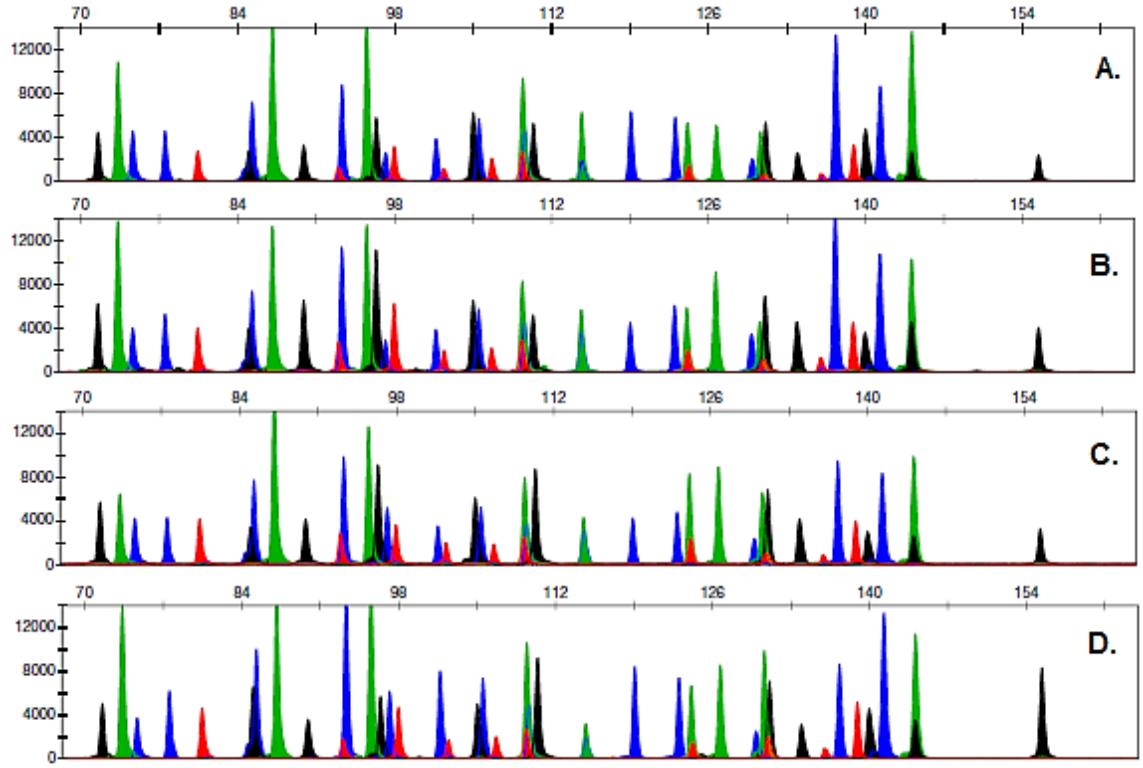


Figure 3-1: The sensitivity of the INDEL kit used in this research. Four different amount of the PCR mixed; A) one half; B) one third; C) one fourth and D) one fifth of the recommended volume of 25µl were tested. The master mixed was used on (2.5, 1.6, 1.25 and 1.0) µl respectively of Control DNA 9948 (500 pg/µl). The results showed that optimum profile still could be obtained in one fifth of PCR reaction volume.

3.4.2 Evaluation of DNA purification methods

Two different purification methods were used to extract DNA from blood stained on FTA card. These methods were carried out on ten samples. Then, PCR amplification performed in final volume of one fifth of the manufacturer's recommendation (see Figure 3-2). The results showed allele dropouts were obtained with direct PCR on purified FTA card. Whereas, samples which were extracted using organic method followed by purification using Microcon® Centrifugal Filter Devices (Merck Milipore), showed more reliable results. No PCR amplification was detected in the negative control, indicating absence of contamination**.**

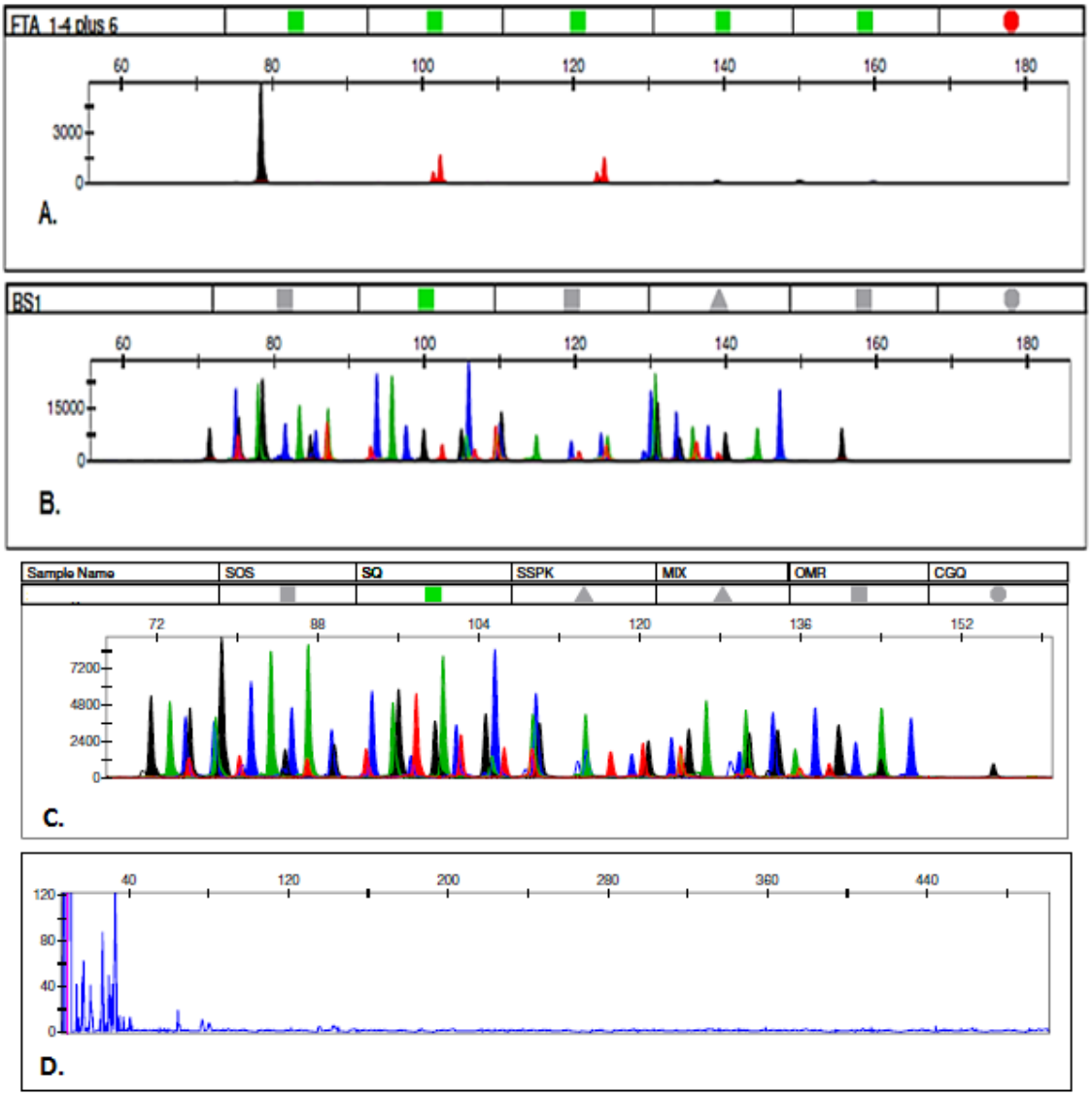


Figure 3-2: The example of electropherograms generated from A) direct FTA punch, washed prior amplification and B) FTA card extracted using phenol chloroform, quantitated (the concentration ranged from 4.0 ng/µl to 10.0 ng/µl) and amplification (refer to 3.3). C) positive control and D) negative control. No DNA profile was developed at negative control

3.4.3 Volume of Allelic Ladder.

A volume of 0.5 µl and 1.0 µl of allelic ladder Qiagen Investigator DIPplex® kit was loaded with each batch of samples which prepared for capillary electrophoresis. A cocktail consisting of size standard and HiDi formamide was prepared by adding 0.5 µl GeneScan™ 500 LIZ™ dye Size Standard (Life Technologies) and 8.5 µl of HiDI formamide together, per sample. A volume of 9.0 µl of the cocktail was added to the 1 µl allelic ladder. There were no differences in the results obtained.

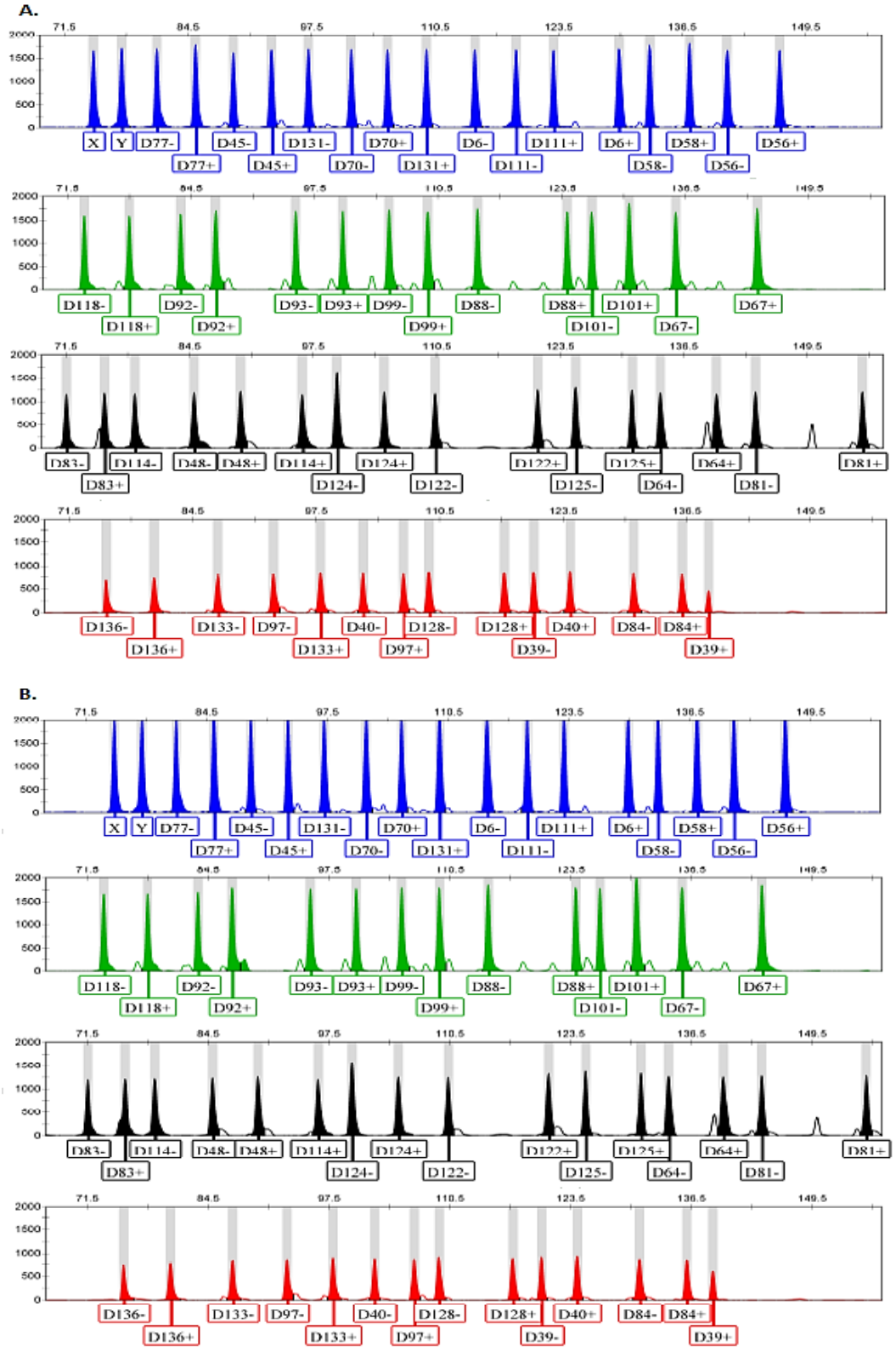


Figure 3-3: Electropherograms of all 62 alleles present in the allelic ladder A) 0.5 µl and B) 1.0 µl, in 9.0 µl of mixture (0.5 µl GeneScan™ 500 LIZ™ dye Size Standard and 8.5 µl of HiDi formamide).

3.4.5 Use of size standard Genescan™ 500 LIZ®

Five samples were run using the size standard 550 (BTO) that was provided in the Investigator DIPplex® kit (Qiagen). Whilst another five samples were run using Genescan™ 500 LIZ® Size Standard (Applied Biosystems). The reaction volume which was used: 8.5 µl of Hi-Di™ Formamide and 0.5 µl of size standard. However, the size standard 550 (BTO) often was not defined or identified correctly. As a result, either off-ladder profiles were produced or the DNA profiles could not be analysed. Hence, size standard Genescan™ 500 LIZ® was proved to be more robust, as sizes being correctly identified more often compared to size standard 550 (BTO). However, the filter set remains the same (i.e., G5) and the BT5 matrix set is used (Bashir, Hassan 2016).

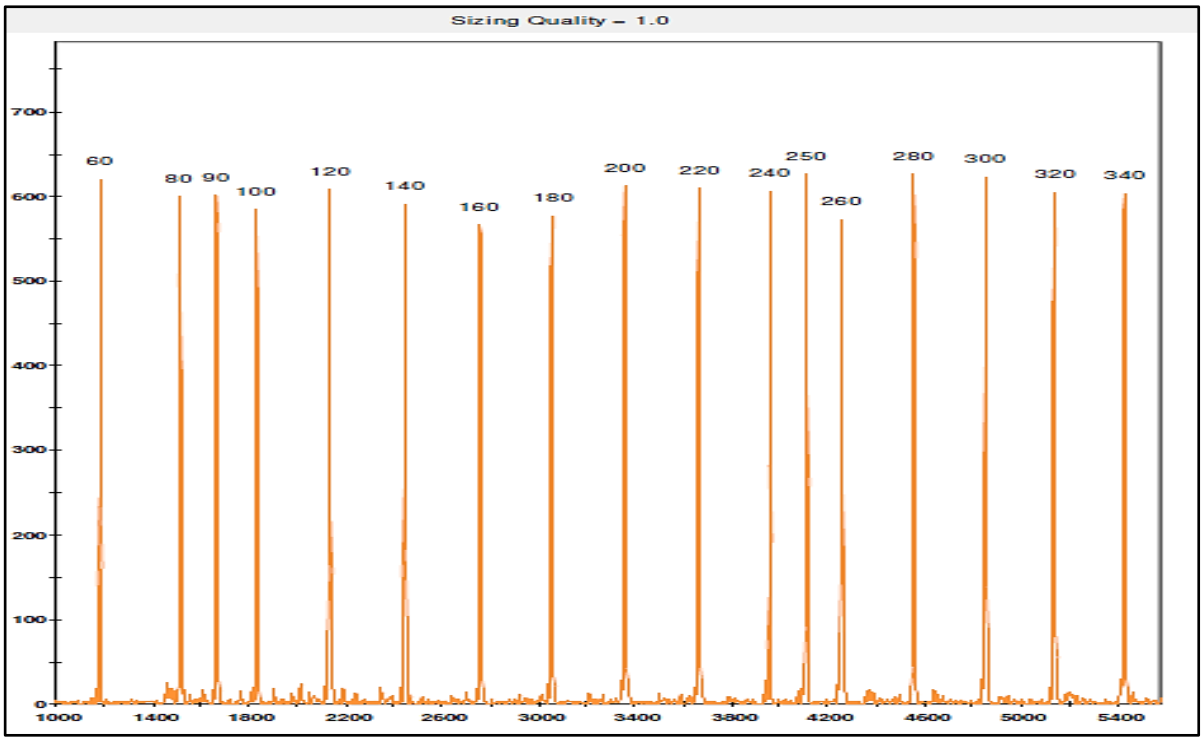


Figure 3-4: The example of electropherogram of size standard 550 (BTO). The sizing ranged 60 bp to 340 bp and height were around 570 rfu to 650 rfu.

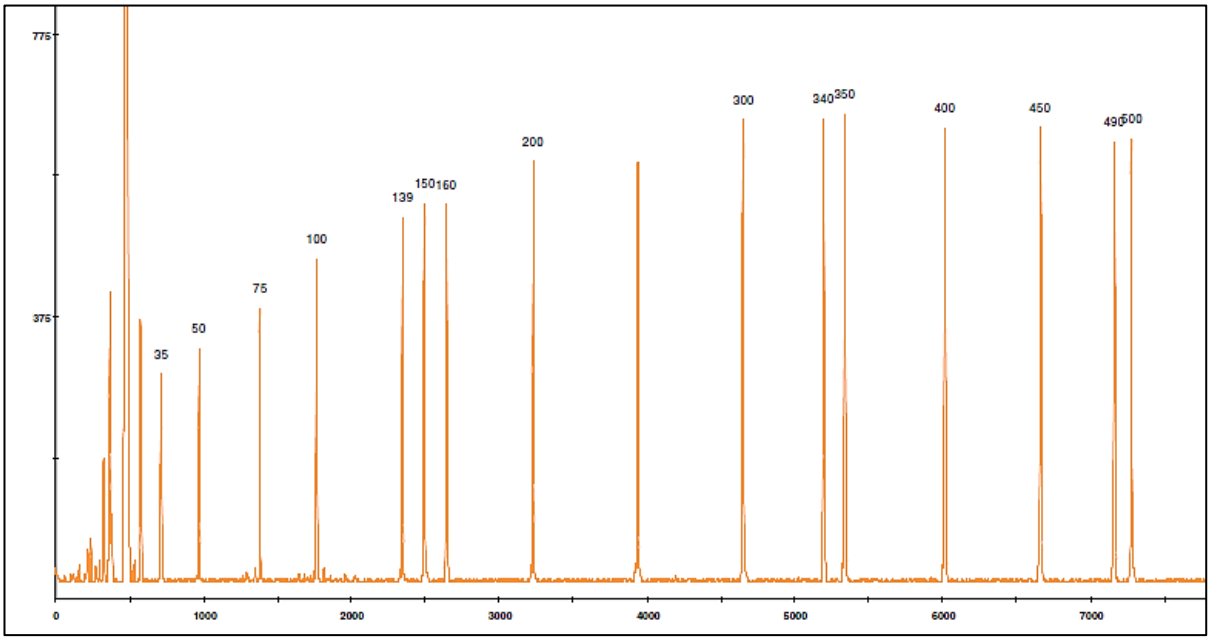


Figure 3-5: The example of electropherogram of size standard Genescan™ 500 LIZ®. The sizing ranged 50 bp to 500 bp and the height were around 1600 rfu to 3200 rfu.

3.4.6 Sensitivity study

A sensitivity study of the assay is essential to evaluate the ability of generating reliable profiles from a range of DNA quantities. Serial dilution of the Human Genome Female Control (Promega) with DNA quantities of at (10, 5, 2.5, 1.25, 0.6, 0.3, 0.15, 0.07, 0.03, 0.01 and 0.009) ng. All samples were successfully amplified but, partial profiles were amplified from as low as 0.07 ng of the control DNA dilutions (see Table 3-3). This shows the result were robust down to 0.15 ng. Input DNA of 0.15 ng to 0.6 ng resulted in DNA profiles of better quality consisting of amplification of all 30 loci and good peak morphology (see Figure 3-6). A high concentration of input DNA (1.25 ng and higher) resulted in off scale peaks, pull ups, split peaks and in some cases, allele drop in due to the peaks being above the detection level of the 3500 Genetic Analyzer (Life Technologies) as well as massive peak imbalance. PCR amplification using input DNA of 0.03 ng and below proved to be insufficient, resulting in DNA profiles of poor quality and incomplete profiles with allelic dropout.

Table 3-3: Investigator DIPplex® kit allele calls for Female DNA Control (Promega) for the two template range; 10 ng to 0.009 ng.

Locus	DNA concentrations (ng/µl)										
	10	5	2.5	1.25	0.6	0.3	0.15	0.07	0.03	0.01	0.009
Amel	X/X	X/X	X/X	X/X	X/X	X/X	X/X	X/X	X/X	X/X	X/X
HLD77	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+
HLD45	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+
HLD131	-/*	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+
HLD70	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+
HLD6	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/*	*/+	-/*
HLD111	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/*	-/+
HLD58	-/+	-/+	-/+	-/+	-/+	-/+	-/+	*/+	-/+	-/+	-/+
HLD56	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/*
HLD118	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+
HLD92	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	*/*
HLD93	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+
HLD99	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/*	-/*	-/+
HLD88	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	*/+	-/+
HLD101	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+
HLD67	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/*	*/*	*/*
HLD83	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/*	-/+	-/+	-/*
HLD114	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	*/*
HLD48	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	*/+
HLD124	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+
HLD122	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+
HLD125	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+
HLD64	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/*
HLD81	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	*/+
HLD136	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+
HLD133	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/*	*/*
HLD97	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+
HLD40	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	*/+
HLD128	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	*/*
HLD39	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/*	*/+	-/*
HLD84	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+	-/+

‘-‘ represents deletion and ‘+’ represents insertion; * represents allele drop out.

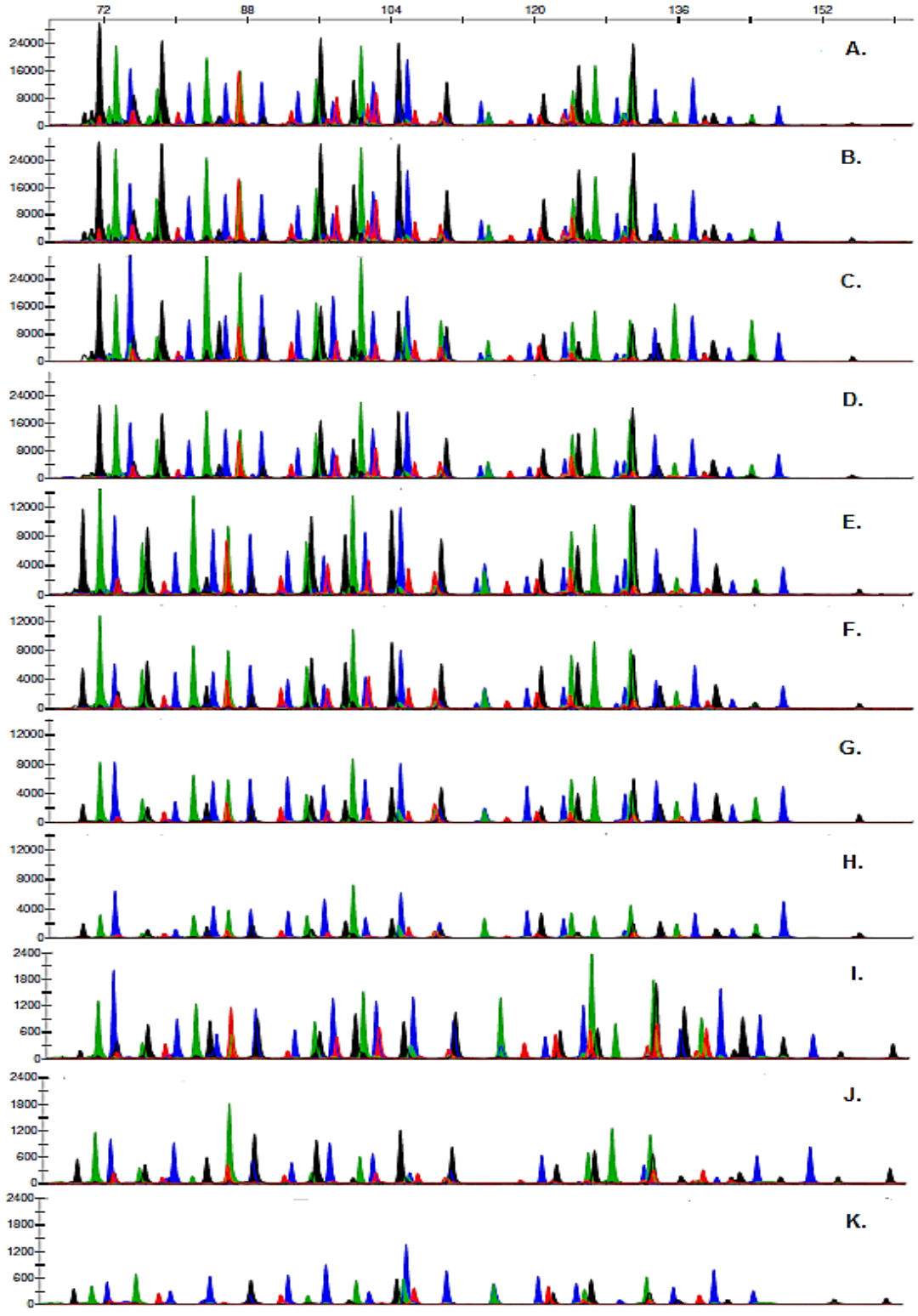


Figure 3-6: INDEL profiles generated from a series of dilution of Human Genome Female Control (Promega) with concentrations of (A) 10.0 ng/µl, (B) 5.0 ng/µl, (C) 2.5 ng/µl (D) 1.25 ng/µl, (E) 0.6 ng/µl, (F) 0.3 ng/µl, (G) 0.15 ng/µl (H) 0.07 ng/µl, (I) 0.03 ng/µl, (J) 0.01 ng/µl and (K) 0.009 ng/µl. However, this assay able to type DNA and obtain full genotypes with 0.15 ng of template DNA. Different RFU scales y-axis are shown to demonstate degradation.

3.4.6.1 Intra-colour balance

The intra colour peak height balance in each fluorescent dye channel was calculated for the profile analysed using the Investigator DIPplex® Kit to find out the performance of dyes. Regardless to the DNA concentration, DIPplex generated good quality of INDEL profiles, producing well-balanced peak heights within each dye channel. The DIPplex® profiles were observed to exhibit intra colour balance (ICB) within the range of 14 to 100 % across the profiles. This shows that the dyes are still sensitive even with a low template of DNA. Ideally, the ICB should be 40% to 50% indicated marker-to-marker colour within a dye chanel. When the total number of templates added to the PCR is extremely low, unbalanced amplification of the alleles may occur due to stochastic effects.

Table 3-4: The intra colour balance of Investigator DIPplex® Kit in 5 µl of PCR reaction.

DNA template (ng/µl)	Intra colour balance (%)			
	DIPplex Blue (6-FAM)	DIPplex Green (BTG)	DIPplex Yellow (BTY)	DIPplex Red (BTR)
1.0	30	33	14	19
0.5	38	43	26	33
0.25	45	45	41	50
0.125	40	30	38	41
0.06	55	35	39	24
0.03	44	24	43	47
0.015	96	60	93	100

3.4.6.2 Peak height ratio

The peak height ratio for the heterozygous loci in Human Genomic Female DNA is illustrated in Table 3-5. The mean heterozygote ratio for DNA template range of 0.015 to 1.0 ng in 5.0 µl of PCR reaction volume exceeded 20% for all markers.

Table 3-5: Peak height ratio of Investigator DIPplex® Kit for Human Genomic Female DNA in 5.0 µl per reaction volume

INDEL Locus	1.0 ng	0.5 ng	0.25 ng	0.125 ng	0.06 ng	0.03 ng	0.015 ng
	Peak height ratio (%)						
Amel	92	91	77	70	60	83	57
HLD77	73	78	84	71	65	86	AD
HLD45	56	59	65	56	52	66	AD
HLD131	26	30	28	34	25	AD	AD
HLD70	41	46	38	37	47	74	AD
HLD6	98	70	93	55	57	AD	LD
HLD111	61	65	63	60	99	83	67
HLD58	94	76	84	76	44	90	AD
HLD56	60	52	45	66	98	97	AD
HLD118	80	85	68	59	88	30	91
HLD92	95	93	88	58	61	67	87
HLD93	61	63	74	60	65	67	90
HLD99	34	30	30	48	60	51	AD
HLD88	40	50	64	47	51	69	AD
HLD101	87	86	93	98	64	35	LD
HLD67	41	39	36	35	90	33	67
HLD83	85	84	77	98	54	97	AD
HLD114	63	69	63	65	31	79	AD
HLD48	85	71	97	72	96	62	37
HLD124	89	68	95	72	66	82	75
HLD122	67	74	80	77	78	33	AD
HLD125	93	68	78	32	29	83	AD
HLD64	90	93	90	80	82	79	LD
HLD81	80	89	65	84	91	85	AD
HLD136	90	85	97	70	20	AD	LD
HLD133	23	21	22	12	AD	36	LD
HLD97	96	93	77	71	39	47	LD
HLD40	75	92	69	49	90	30	LD
HLD128	89	91	92	79	81	45	LD
HLD39	43	56	45	52	25	64	LD
HLD84	95	72	71	70	61	74	100

AD represents Allele dropout; LD represents Locus droput

3.4.7 PCR cycles

The standard 30 cycles which recommended by the manufacturer worked well with the low template range to 0.07 ng to 0.6 ng DNA. The lower template DNA samples; I, J and K with 0.03, 0.01 and 0.009 ng respectively gave inconclusive result were amplified using 32 cycles. These samples were tested in duplicate. Full profile was obtained at 0.03 ng and partial profiles at 0.01 and 0.009 ng respectively. However, with 32 cycles (see Figure 3-7) caution should be exercised as extra peaks or allele dropout could be detected which possibly lead to false interpretation.

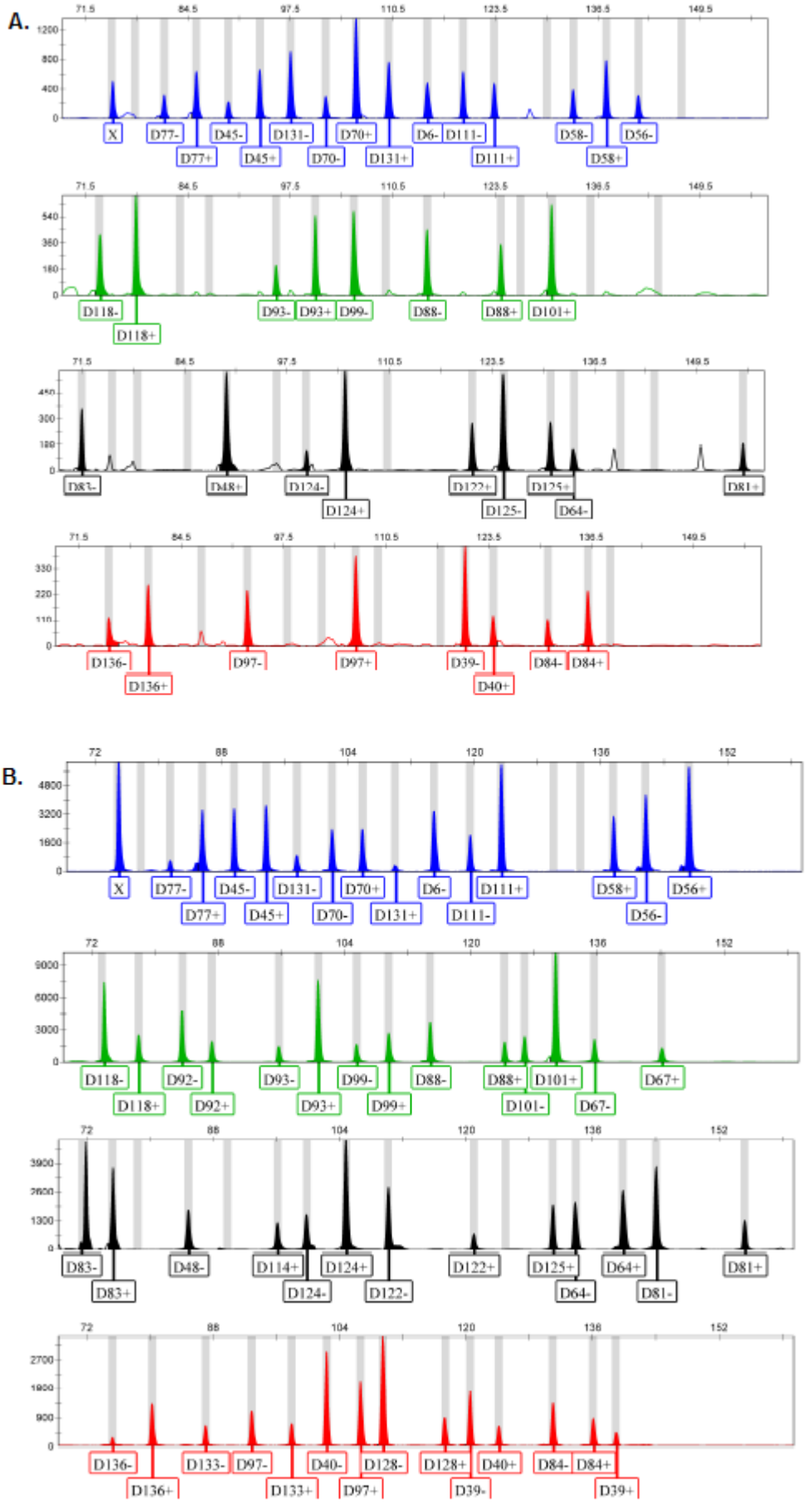

A.
71.5
84.5
97.5
110.5
123.5
136.5
149.5
X
D77-
D77+
D45-
D45+
D131-
D70-
D70+
D131+
D6-
D111-
D111+
D58-
D58+
D56-
D118-
D118+
D93-
D93+
D99-
D88-
D88+
D101+
D83-
D48+
D124-
D124+
D122+
D125-
D125+
D64-
D81+
D136-
D136+
D97-
D97+
D39-
D40+
D84-
D84+
B.
72
88
104
120
136
152
X
D77-
D77+
D45-
D45+
D131-
D70-
D70+
D131+
D6-
D111-
D111+
D58+
D56-
D56+
D118-
D118+
D92-
D92+
D93-
D93+
D99-
D99+
D88-
D88+
D101-
D101+
D67-
D67+
D83-
D83+
D48-
D114+
D124-
D124+
D122-
D122+
D125+
D64-
D64+
D81-
D81+
D136-
D136+
D133-
D133+
D97-
D97+
D40-
D128-
D128+
D39-
D40+
D84-
D84+
D39+

Figure 3-7: A) A DIPplex® profile from the diluted samples using the human Female DNA Control (Promega) and amplified for 30 cycles, with 0.009 ng/µl of input DNA in 5.0 µl reaction volume, B) The profile with 30 markers and is produced a partial profile and unacceptable of dropouts compared to 32 cycles. Allele Y (purple arrow) at (B) can be seen and the RFU is above 50. This can be mistakenly identified as real peak. Also, there were allele drop ins (red arrow) observed at 32 cycles.

3.4.8 DNA Profile quality

Lower concentration of DNA; diluted Human Genomic Female DNA (Promega) ranging from 1.0 ng to 0.015 ng were used to see the peak height of the allele obtained. When used in lower PCR volumes, 5 µl, the DIPplex® produced significantly higher (Leclair et al. 2003) peak heights than the peak height set by the manufacturer. Heterozygote peak imbalance was observed across all seven concentrations tested, indicates that the pair of alleles associated with each heterozygous locus was relatively balanced especially at 0.125 ng up to 1 ng. Whereas, in samples using DNA template in the range of 0.015 ng to 0.06 ng, the heterozygote peak imbalances were observed however it was not high enough to impact allele calling and interpretation.

3.4.9 Artificial DNA degradation study

In order to assess the performance of the assay in cases of poor quality DNA, a degradation study was performed. Environmental insults on forensic samples may result in DNA degradation or damage at random locations. As with any multi-locus system, the possibility exists that not every locus will amplify if the DNA sample has been severely degraded with the largest sized loci being the most susceptible.

The ability of the Investigator DIPplex® kit to amplify degraded DNA was investigated. Artificially degraded DNA samples were prepared by incubating purified genomic DNA digestion with Promega RQ1 RNase-Free DNase I (100x diluted) and serial degradation were taken at intervals from 2 min to 180 min. One nanogram of degraded DNA was amplified using the Investigator DIPplex® kit. Longer incubation times lead to further degradation and smaller average fragments. With increasing degradation, the signal intensity of the INDEL loci decreased equally over the full size range with only slightly stronger decrease for the longer amplicon lengths as shown in Figure 3.8.

Full profiles were detected for degradation times of up to 10 min, with only 11 alleles of 61 alleles dropping out after 20 mins of incubation. Even after 30 mins, it was still possible to detect 41 of the 60 possible alleles, although at very low signal intensities. More allele

dropout as the DNA became increasingly degraded. Preferential amplification was not observed but the loci able to robustly amplify in the order of decreasing size as the extent of degradation progressed.

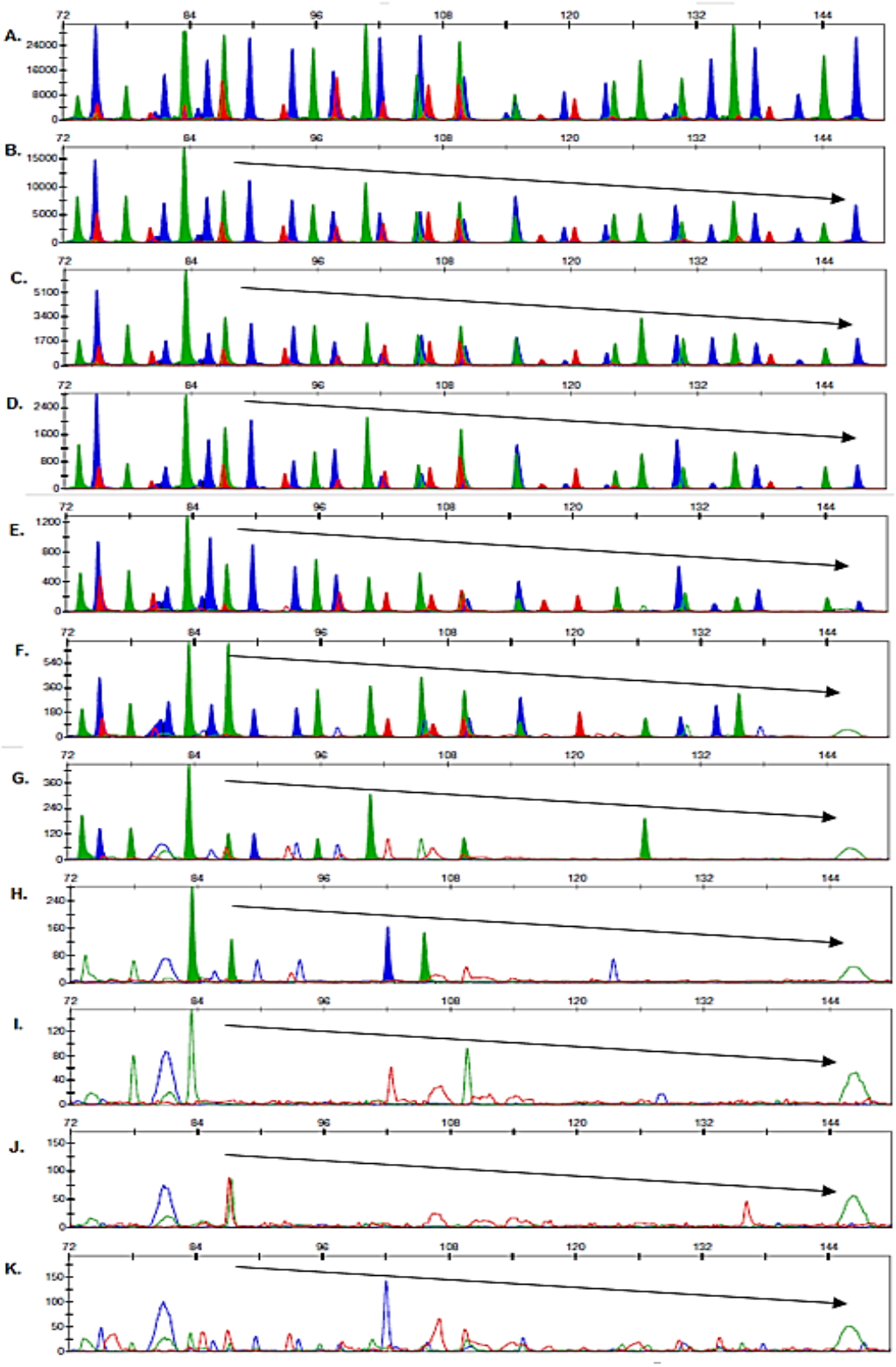


Figure 3-8: Electropherogram shows DIPplex® amplification of artificially degraded DNA using DNase I (100x diluted) on Human Genomic Female DNA at different time of points (A) 0, (B) 2, (C) 5, (D) 10, (E) 20, (F) 30, (G) 45, (H) 60, (I) 90, (J) 120 and (K) 180 min. Different RFU scales y-axis are shown to demonstrate degradation.

3.4 9 The INDEL multiplex performance

3.4.9.1 Malaysian population samples

Amplification of samples for the purpose of constructing a reference database was performed using the amplification protocol consisting of a final volume of 5 µl and thermal cycling conditions as recommended by the manufacturer. Optimal DNA concentration for input DNA was determined to be between 0.2 to 0.5 ng. PCR amplification was performed using input DNA of 0.5 ng for samples extracted from blood stained specimen.

A total of 500 blood stained specimen were collected and DNA was successfully extracted from the samples (100 each of the Malays, Chinese, M-Indian, Iban and Bidayuh). The DNA phenol-chloroform method proved to be efficient, with 90% of the stock DNA having concentrations of more than 50 ng/µl. Amplified fragments, run on the ABI 3500 instrument, typically generated profiles that could be easily interpreted. Examples of such electropherogram for the Qiagen Investigator DIPplex® kit reaction for the Malays, Chinese, M-Indian, Iban and Bidayuh are presented in Figure 3-9, 3-10, 3-11, 3-12 and 3-13 respectively.

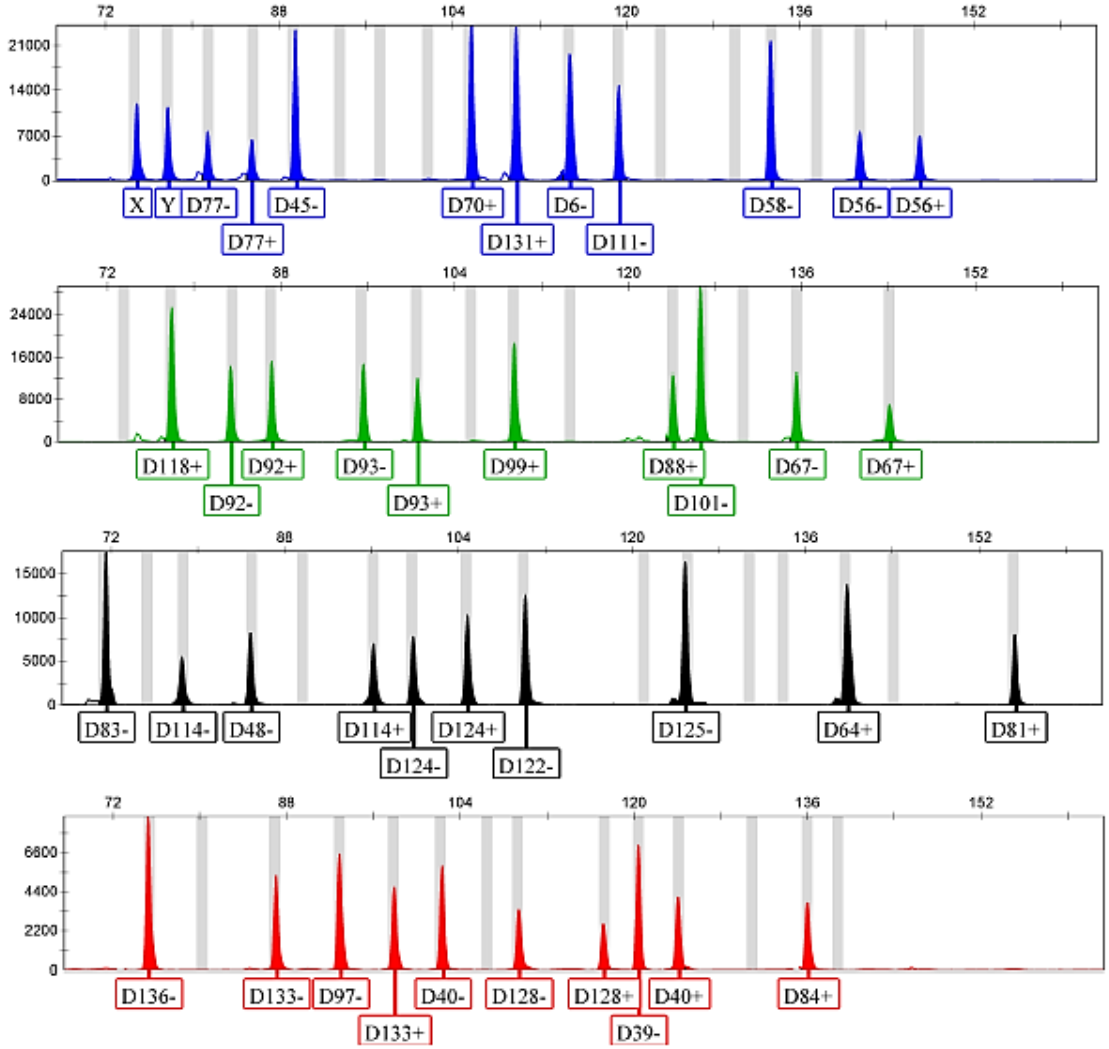


Figure 3-9: An example profile of the Malays population profile using Qiagen Investigator® DIPplex Kit.

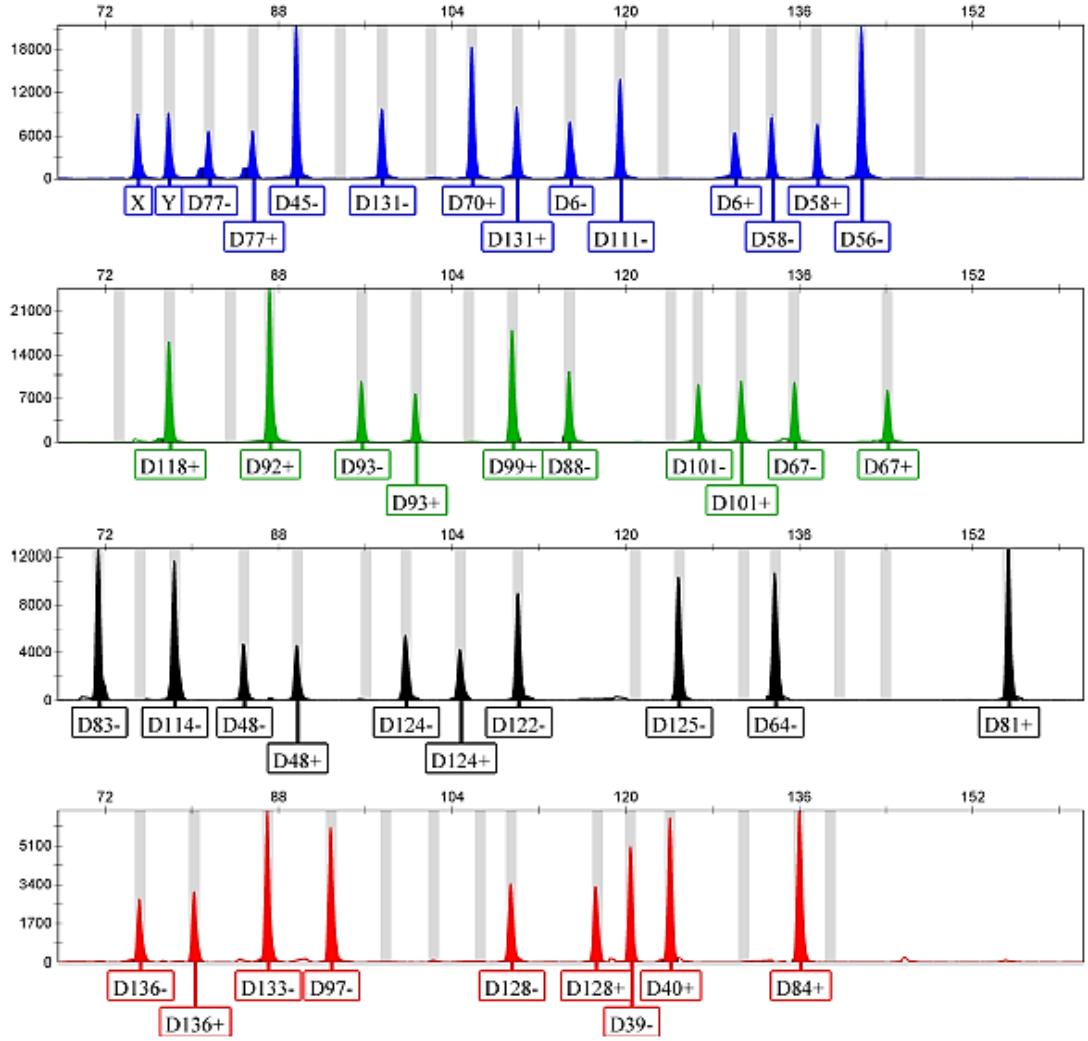


Figure 3-10**:** An example profile of the Chinese population profile using Qiagen Investigator® DIPplex Kit.

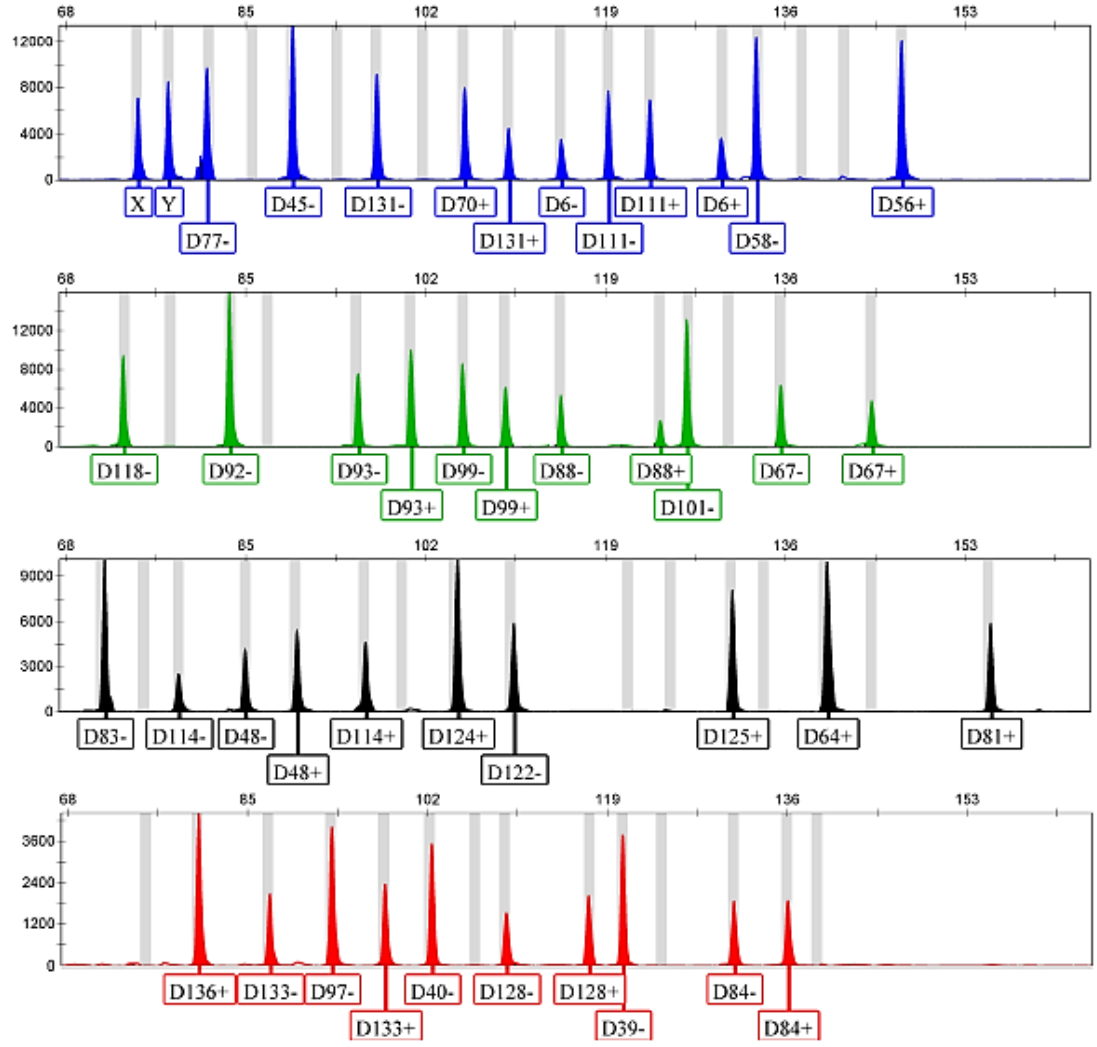


Figure 3-11: An example profile of the M-Indian population profile using Qiagen Investigator® DIPplex Kit.

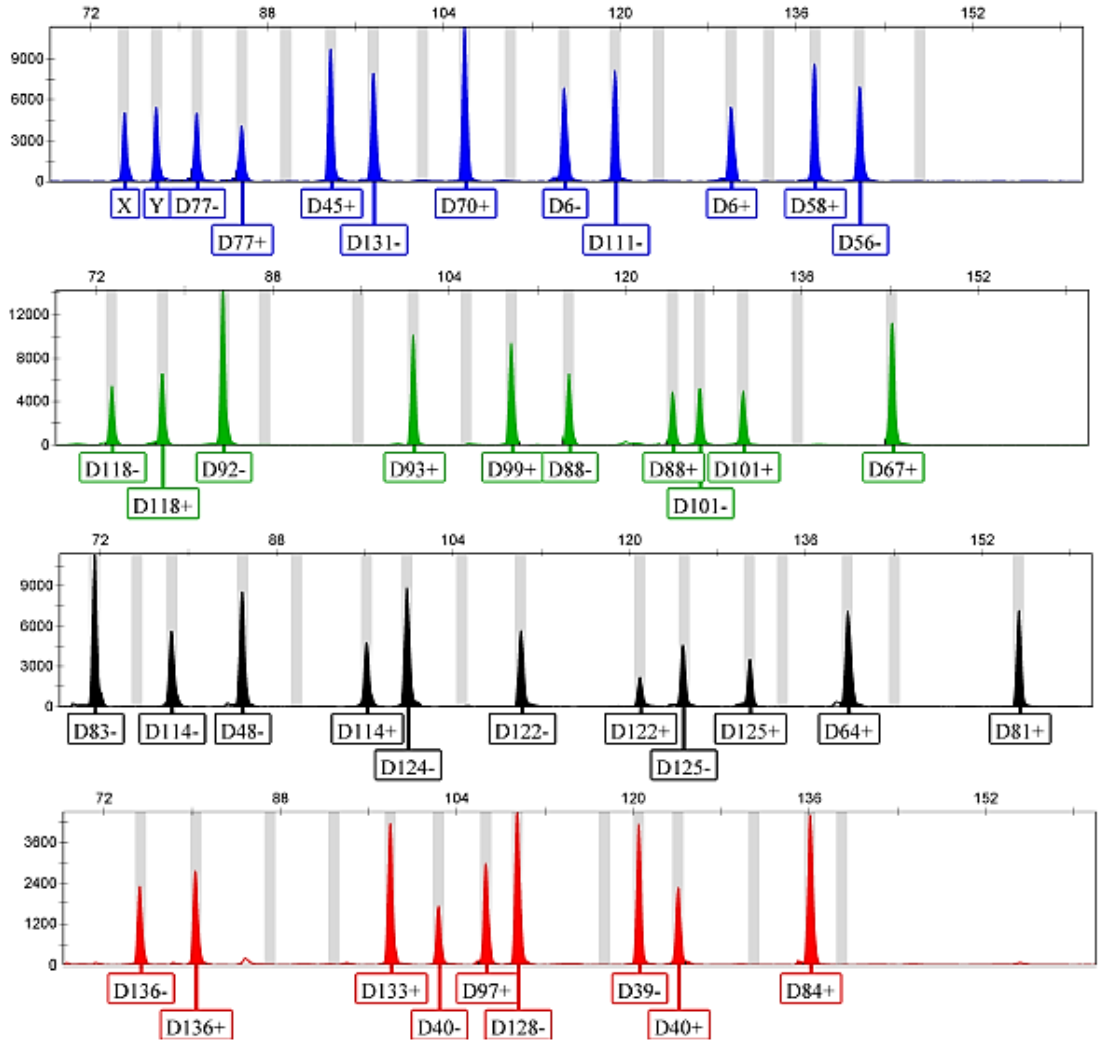


Figure 3-12: An example profile of the Iban population profile using Qiagen Investigator® DIPplex Kit.

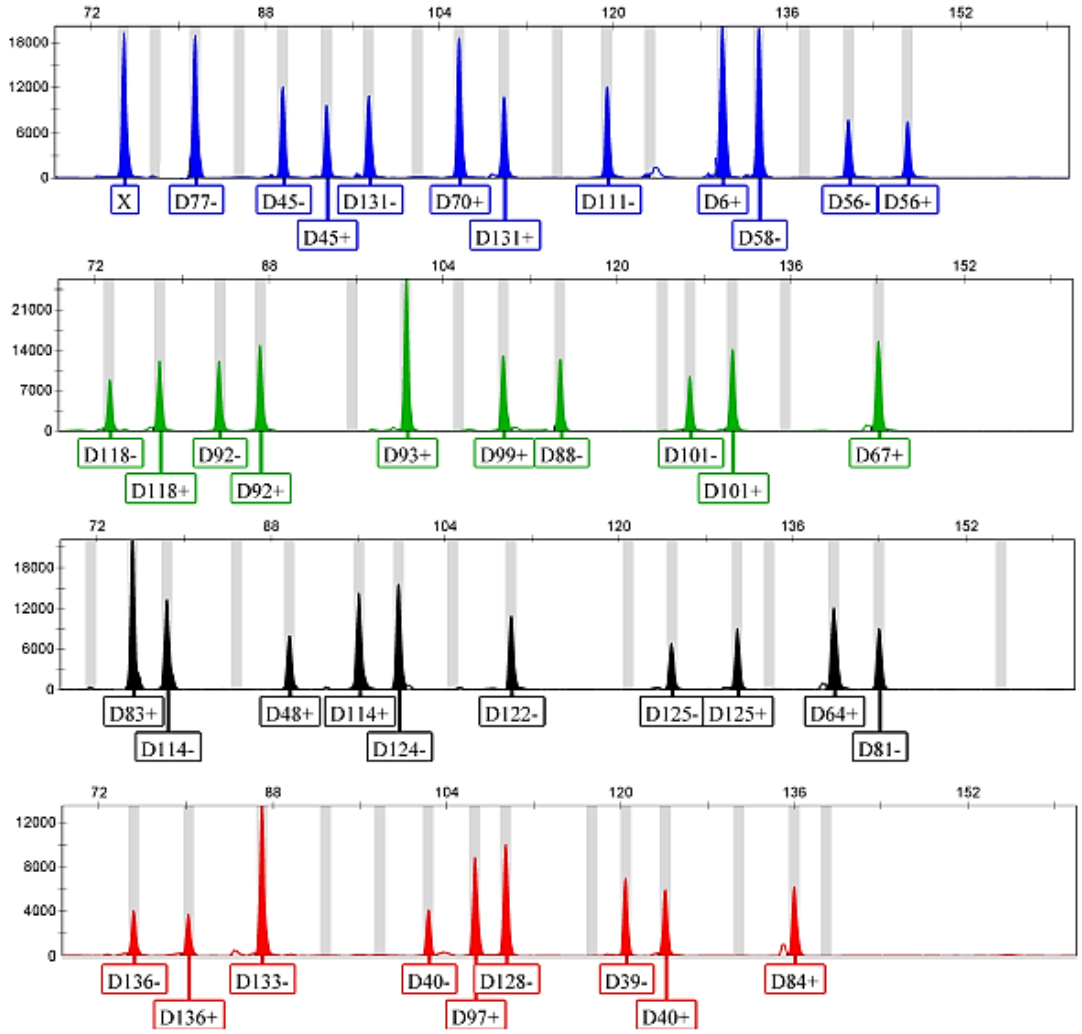


Figure 3-13: An example profile of the Bidayuh population profile using Qiagen Investigator® DIPplex Kit.

3.4.9.2 Forensic Samples

DNA concentration from various challenged samples ranged from undetected amount to 40 ng/µl. In this study, bloodstain and bone samples which accounted for most of the samples with concentrations of DNA below detection level. Some of these samples yielded either partial or no STR profiles. It was possible to generate complete and concordant DIPplex® profile from all of these samples.

3.4.9.3 Crime samples

Twenty-eight crime samples comprising bloodstains on pillow case, bedsheet, blanket, skirt, comforter, jeans, blouse, camisole, shirt, towel, jeans, underwear, short pants and singlet from eight murder cases were profiled using Powerplex® 16 System and Investigator DIPplex® kit. Results obtained from both analyses were shown in Table 3-6 respectively.

Out of 28 selected samples; 3 full STR profile were generated using conventional STR analysis; of the 25 samples 4 gave partial profiles and 21 could not be amplified at all. In comparison, with 30 INDELS markers, 11 full profiles and only 17 were partially amplified or not amplified at all (with 8 partial profiles and no DNA profile were obtained from 9 samples).

Figure 3-14, the DNA partial STR and full INDEL profiles were generated from the crime samples; the DNA concentration was 0.22 ng/µl. In Figure 3-15, no STR loci was amplified and partial INDEL profiles were generated from the crime samples. The DNA concentration was 0.001 ng/µl.

Table 3-6: Table below shows in house label of samples, origin of samples and the results obtained from analysis of 15 STRs and 30 INDELS. MP represents match probability. The MP (match probability) of the INDEL loci were also calculated.

Case	Label	Samples	DNA concentration (ng/µl)	STR	MP	INDEL	MP
1	F1	Bloodstains on pillow case	0.24	---		---	
	F2	Bloodstains on bed sheet	0.61	---		---	
	F3	Bloodstains on blanket	40.1	---		---	
2	F4	Bloodstains on skirt	0.12	++- [10 loci]	3.42×10^{-16}	+++	5.56×10^{-13}
3	F5	Bloodstains on comforter 1	0.001	---		---	9.64×10^{-13}
	F6	Bloodstains on comforter 2	0.05	++- [13 loci]	1.58×10^{14}	+++	
4	F7	Bloodstains on a pair of jeans	0.37	+++	6.55×10^{17}	+++	2.39×10^{-12}
	F8	Bloodstains on blouse	0.001	---		---	
	F9	Bloodstains on camisole	Nd	---		---	
5	F10	Bloodstains on blanket 1	0.22	++- [3 loci]		+++	2.39×10^{-12}
	F11	Bloodstains on blanket 2	0.02	+++	7.10×10^{22}	+++	
	F12	Bloodstains on shirt 2	0.89	---		++- [9 loci]	
	F13	Bloodstains on shirt 3	0.01	---		++- [25 loci]	
	F14	Bloodstains on shirt 4	0.001	---		---	
	F15	Bloodstains on shirt 5	0.03	---		+++	
	F16	Bloodstains on shirt 6	Nd	---		++- [24 loci]	
	F17	Bloodstains on shirt 7	Nd	---		++- [26 loci]	
	F18	Bloodstains on shirt 8	0.03	---		++- [12 loci]	
6	F19	Bloodstains on a pair of jeans	Nd	---		+++	2.37×10^{-13}
	F20	Bloodstains on towel	Nd	---		---	
	F21	Bloodstains on tissue paper	7.65	++- [11 loci]	1.32×10^{18}	+++	8.33×10^{-13}
7	F22	Bloodstains on T-shirt	0.04	---		++- [26 loci]	4.31×10^{-11}
	F23	Bloodstains on underwear	Nd	---		++- [28 loci]	2.48×10^{-13}
8	F24	Bloodstains on shirt 1	1.05	++-		+++	9.43×10^{-12}
	F25	Bloodstains on shirt 2	5.22	++- [11 loci]		+++	
	F26	Bloodstains on a pair of short pants	1.71	+++	4.87×10^{22}	+++	
	F27	Bloodstains on singlet	0.003	---		++- [18 loci]	
	F28	Bloodstains on a pair of short pants	0.001	---		++- [24 loci	

Note: +++ full profile; ++- partial profile; --- failure profile

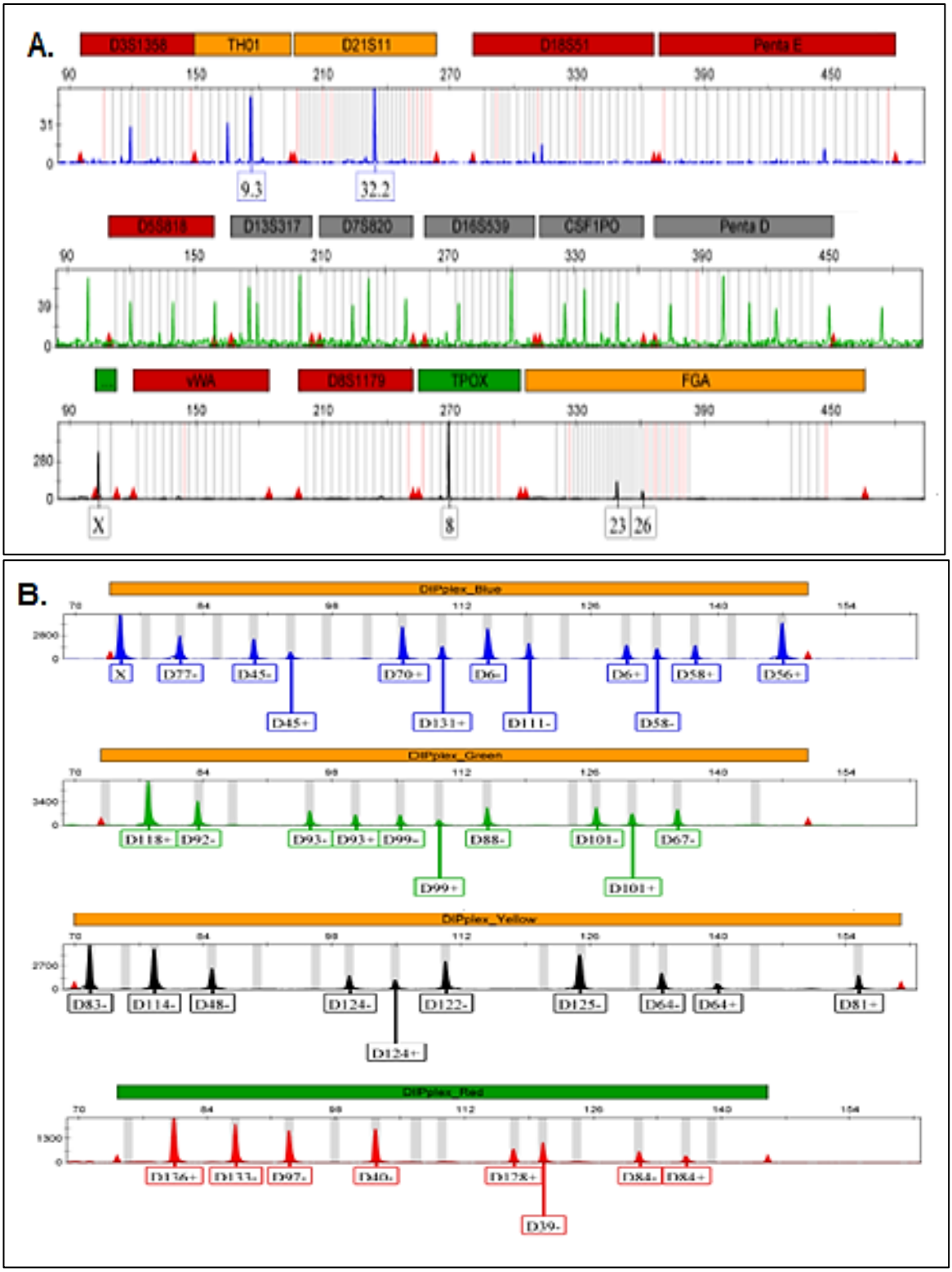


Figure 3-14: Application of STR and INDEL assays on the crime samples with DNA concentrations 0.22 ng/µl. (A) Incomplete DNA profile using STRs but (B) a full INDEL profile generated using Investigator DIPplex® Kit.

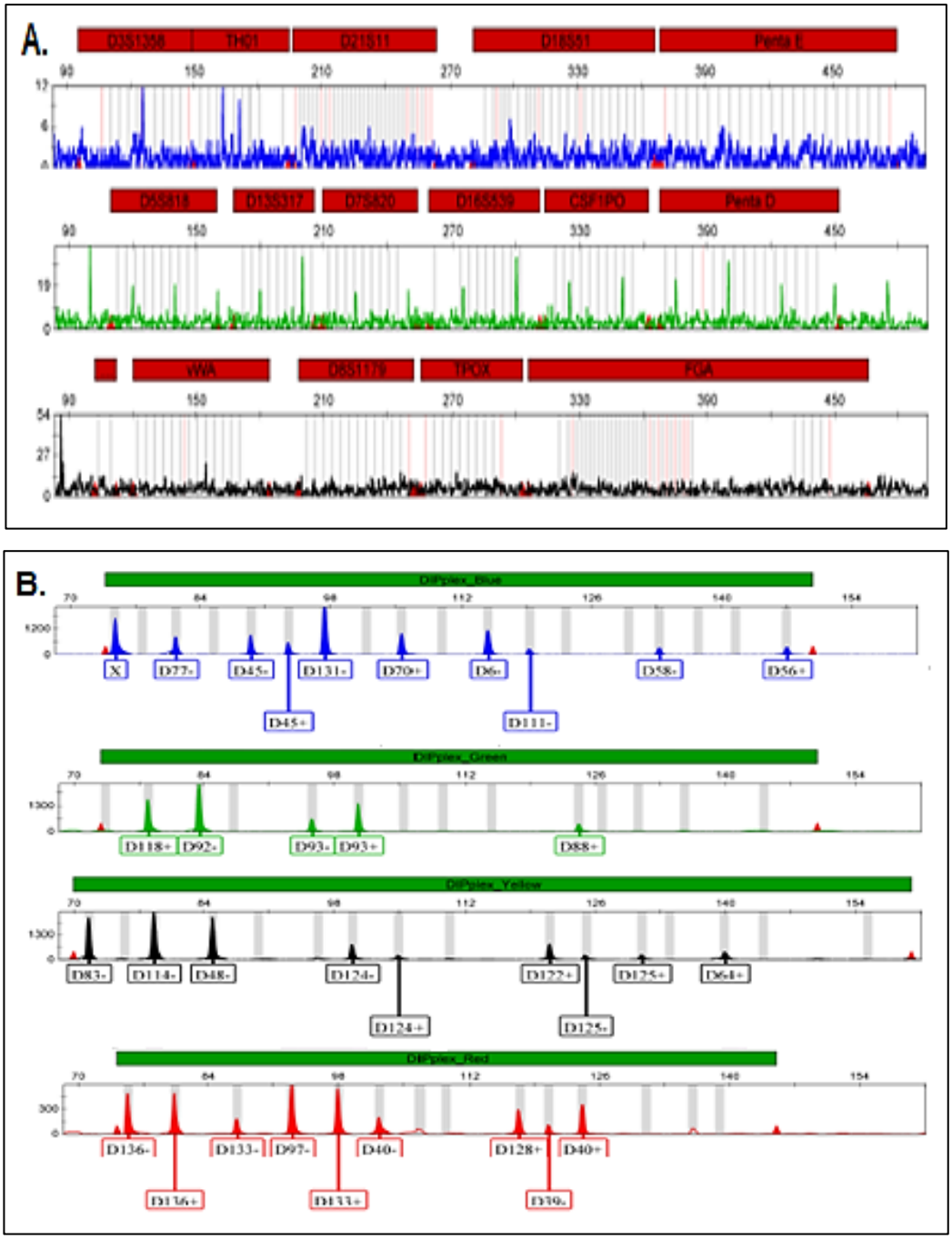


Figure 3-15: Application of STR and INDEL assays on a crime sample with DNA concentration 0.001 ng/µl. (A) No STR loci was amplified but (B) a partial INDEL profile (at 25 loci) generated using Investigator DIPplex® Kit.

3.4.9.4 Bone samples

Thirty bone samples (which were taken from various parts) were profiled using Powerplex® 16 System and Investigator DIPplex® kit. Results obtained from both analyses are shown in Table 3-7.

After DNA quantitation by real time PCR, the concentrations were observed; ranging from 0.000 ng/µl (Nd) to 5.32 ng/µl. Of the 30 bone samples, 6 showed complete INDEL profiles, 6 showed partial profiles and the rest no DNA profiles were detected.

On comparison, amplifying the same forensic samples using STR through Powerplex® kit, 4 gave partial STR profiles and for the remaining samples, no DNA profiles were detected on the bone samples.

The examples of results obtained from the bone samples are shown in Figure 3-16 and Figure 3-17. In Figure 3-16, the DNA partial STR and full INDEL profiles were generated from the bone samples; the DNA concentration was 0.70 ng/µl. In Figure 3-17, no STR loci was amplified and partial INDEL profiles were generated from the bone samples; the DNA concentration was 0.007 ng/µl.

Table 3-7: The in house label of bones, types of bones and the results obtained from analysis of 15 STRs and 30 INDELS. The MP (match probability) of the INDEL loci were also calculated.

Label	Bone type	DNA concentration (ng/µl)	STR	MP	INDEL	MP
P1	Unknown	0.007	---		++- [17 loci]	1.14×10^{-10}
P2	Unknown	0.37	++- [11 loci]	1.42×10^{14}	+++	2.90×10^{-14}
P3	Unknown	5.32	++- [7 loci]	26100	+++	1.18×10^{-12}
P4	Unknown	Nd	---		---	
P5	Unknown	0.70	+++ [14 loci]	1.58×10^{16}	+++	2.88×10^{-13}
P6	Unknown	0.01	---		++- [14 loci]	9.70×10^{-7}
P7	Unknown	0.14	---		+++	2.88×10^{-13}
P8	Unknown	0.005	---		---	
P9	Unknown	1.37	++- [11 loci]	7.59×10^{14}	++-	2.88×10^{-13}
P10	Femur	0.001	---		---	
P11	Clavicle	Nd	---		---	
P12	Spine	0.009	---		---	
P13	Spine	0.003	---		---	
P14	Rib	Nd	---		---	
P15	Rib	0.04	---		---	
P16	Rib	Nd	---		---	
P17	Rib	0.05	---		++- [9 loci]	1.23×10^{-6}
P18	Rib	Nd	---		---	
P19	Rib	0.007	---		---	
P20	Rib	0.02	---		++- [17 loci]	1.02×10^{-8}
P21	Rib	0.01	---		---	
P22	Rib	0.09	---		---	
P23	Rib	0.47	---		+++	1.85×10^{-16}
P24	Rib	0.08	---		---	
P25	Rib	0.03	---		---	
P26	Rib	0.05	---		---	
P27	Rib	0.001	---		---	
P28	Rib	0.005	---		---	
P29	Rib	0.05	---		++- [27 loci]	1.94×10^{-16}
P30	Rib	0.008	---		++- [7 loci]	1.03×10^{-3}

Note: +++ full profile; ++- partial profile; --- failure profile

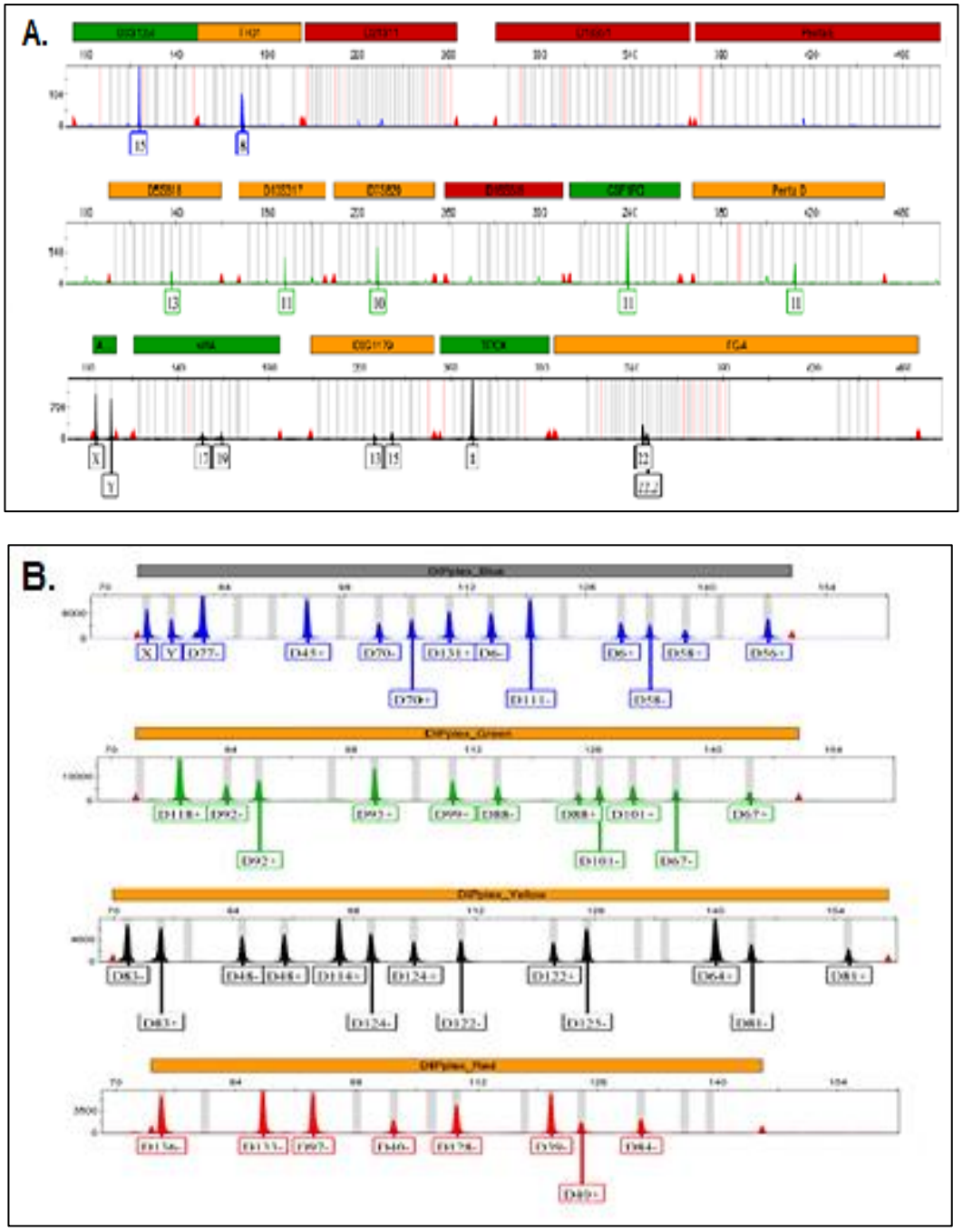


Figure 3-16: Application of STR and INDEL assays on the bone samples with DNA concentration 0.70 ng/µl. (A) Incomplete DNA profile using STRs but (B) a full INDEL profile generated using Investigator DIPplex® Kit.

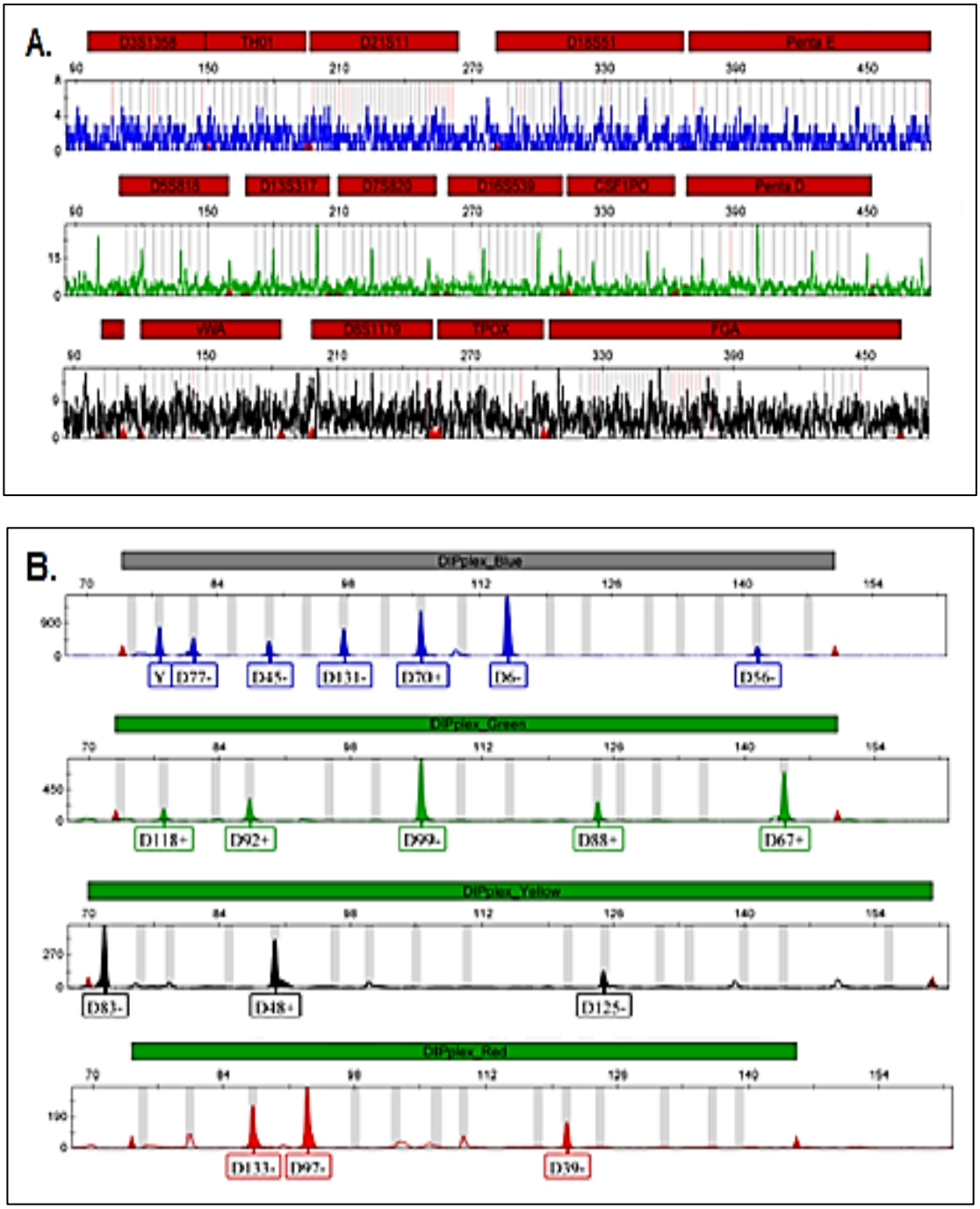


Figure 3-17: Application of STR and INDEL assays on the bone samples with DNA concentrations 0.007 ng/µl. (A) Incomplete DNA profiles using STRs (at 18 loci including at ameloginen) but (B) a full INDEL profile generated using Investigator DIPplex® Kit.

3.5 Discussion

The kinds of samples usually encountered in forensic are degraded and low template DNA samples. So, any PCR system utilized for forensic purposes it is more significance to perform optimisation, including the optimisation of reagents and instrumentation. Optimisation also means optimising the laboratory environment in which the work will be done. Again, due to the sensitivity of the PCR system used in forensics, the laboratory setup has a big impact on the quality of the work. Even these experiments are not exhaustive, but they are appropriate for the Investigator DIPplex® Kit to be applied in forensic and/or parentage testing.

3.5.1 Evaluation of DNA purification methods

The FTA cards of the Malaysian population were kept for more than 10 years in KIMIA DNA Forensic Laboratory. The colour of the blood stained were dark red and the staining was a bit thick. In addition, with the presence of heme compound (Akane et al. 1994, Hu et al. 2015), it is not possible to carry out direct PCR on the FTA card before purification. It has been reported by (Koons et al. 1994), that more DNA yield can be increased using phenol-chloroform method followed by dialysis and concentration with Centricon 100® devices rather than ethanol precipitation. High yield DNA successfully obtained from the reference samples and good quality profiles were developed without any major impact on the peak balance (Bashir, Hassan 2016).

3.5.2 Instrument parameters

The injection and run time recommended by the manufacturer was 8 s and 900 s respectively using POP 4 and 36 cm capillary (Investigator DIPplex Handbook 2010). However, the 3500 Genetic Analyzer at UCLAN laboratory is using POP 6 and 50 cm capillary. Due to the different concentration and viscosity between the polymers, the injection time was increased from 8 s to 10 s whilst run time was extended to 2500 s to increase the sensitivity (Westen et al., 2009).

Other than that, the bins settings needed to be modify with reference to the allelic ladder to compensate for variation, due to the electrophoretic mobility of the fragments on different platforms and under different conditions e.g. environmental temperature (Bashir, Hassan 2016).

3.5.3. Specificity

Non-human DNA samples were not tested in this study. However, a validation study by (Klein et al., 2015) had demonstrated the human specificity of the Qiagen Investigator DIPplex® system. The study which involved tests on a variety of animal species showed partially detectable profiles were derived from primate DNA samples (such as lowland gorilla and brown headed spider monkey) but none of these primates yielded complete profiles as observed with pristine, degraded and challenged samples from humans. Whilst, the DNA samples from non-primate species did not yield reproducible detectable PCR products.

3.5.4 Sensitivity

DNA quality and concentration of input DNA had a big effect on the DNA profiles obtained. The fine balance of the system was demonstrated as too much input DNA affected capillary electrophoresis resulting in off scale peaks, pull up and split peaks. The optimal amount of input DNA added to the AmpFlSTR® Identifiler® Plus PCR Amplification Kit between 0.5 ng to 1.0 ng for 28-cycle amplification (Wang et al., 2012). The optimal amount of DNA under standard conditions is 0.2–0.5 ng (DIPplex Handbook). However, in this study the ideal DNA quantity were 0.15 to 0.6 ng in 5 µl PCR reaction. Above 1 ng of template should not be added to Investigator DIPplex® kit due to saturation issues (Larue et al., 2012), as mentioned above.

In the forensic routine cases, we always obtain various amounts of DNA samples, which are ordeal the sensitivity of detection system. To decide the lower limits amount of DNA sample for a full profile with peaks height above 50 RFU, sensitivity studies of the INDELs were performed using Human Genome DNA that were diluted to from 10 ng to 0.009 ng of DNA were amplified and detected. Reproducibly complete and accurate profiles were obtained when input DNA amount down to 0.15 ng. As the decrease in amount of DNA, partial profiles were observed with alleles dropout. Hence, the lower limit of DNA template for this system was sensitive to 0.15 ng. The serial dilutions were also used to conduct the analytical threshold determination experiment.

In STR studies, intra locus balance and peak height ratio were determined to ensure heterozygote were correctly genotyped relative to quantity of DNA and to facilitate mixture interpretations (Collins et al. 2004). (Seah et al. 2016) reported that the intra

colour peak balance exceeded 40% for most profiles and the mean heterozygote ratio exceeded 60% at most DIPplex® loci. The final volume of the PCR used in the research was 20 µl. However, in this study, the results indicated that 1/5 volume (5 µl) yielded good quality profiles without any major impact on the peak balance (Bashir, Hassan 2016), where the peak heights are proportional to the detected PCR products. Even when profiling reference or casework samples, reduced volumes able to produce stable profiles, as long as sufficient template DNA can be added in 1 µl of sample DNA (Bashir, Hassan 2016). The DIPplex® kit produced 100% concordant alleles at all loci when INDEL markers were amplified. INDELs are still new to forensic community. Therefore, no guidelines have published for quality to determine an appropriate minimum peak height threshold and intra colour peak balance for INDELs, but only based on their own results and instruments using low amounts of input DNA.

3.5.5 Increasing PCR cycles

There many studies have been carried out on to analyse low template DNA samples, namely by increasing the number of detected alleles or increase the signal strength to the levels where the peaks are reliable to interpret (Adamowicz et al., 2014) e.g. purification of DNA samples (Schiffner et al., 2005), amplification using reduced volume (Leclair et al. 2003), post PCR purification (Smith, Ballantyne 2007), increased number of PCR cycles (Whitaker et al., 2001) and increased capillary injection settings (Westen et al., 2009). Even though these methods are effective towards improving the DNA profiles and peak heights, caution should be exercised as these also can produced variety of artefacts, including spectral disruptions, increased baseline and stutter values also allele drop-in (Westen et al., 2009).

The Investigator DIPplex® Kit handbook recommended using 32 cycles when working with low DNA template. Three samples I, J and K (refer Figure 3-8) which indicated lower concentrations of DNA (0.03, 0.01 and 0.009 ng respectively), which produced partial profiles at 30 cycles were chosen to amplify at 32 cycles. As expected, unsuitable results were produced especially at 0.009 ng. The profiles were inconsistent and allele drop ins occurred. Therefore, for low template profiling, each samples need to amplify either duplicate or triplicate to obtain reproducibility results and developed consensus profile (Butler, Hill 2010, Word 2010). In addition, all possible precautions should be taken to

minimise sample contamination, such as carried out amplification in clean area also used purified and concentrated DNA samples (Forster et al.,, 2008).

3.5.6 Degraded DNA study

In many situations in forensic casework, if DNA is exposed to environmental contaminants, such as humidity and extreme heat. In order to evaluate the performance of the system in cases of poor quality DNA, a degradation study was carried out. Purified genomic DNA was incubation with DNase I at 24 ºC in a time course manner to create a DNA degradation model. The degraded DNA samples were then genotyped on the ABI 3500 Genetic Analyzer. The results showed that longer incubation times and further degradation would obtained smaller average fragments. The degraded DNA sample at the time point of (90, 120 and 180) min of incubation were analyzed with the 30 INDELs assay, less than 20% of alleles could be detected and still above 50 RFU. The data showed that the Qiagen Investigator DIPplex® could obtain reliable results from degraded DNA.

In this study also, the INDEL markers were applied to the crime and bone samples, in order to evaluate the usefulness of these INDELs in handling problematic samples. The results showed that the INDEL assay returns more genetic information for blood stains and bone samples, compared to the STR assay. Even the match probability of the STR is higher, INDELs still gives an acceptable value for forensic identification.

3.5.7 Amplification of Malaysian population and challenged samples

A total of 500 blood stained specimens of Malaysian population (the Malays, Chinese, M-Indian, Iban and Bidayuh) were successfully amplified using the Investigator DIPplex® Kit. Sufficient and purified DNA were obtained from all the samples, produced good INDEL profiles. Further study on the Malaysian population will be discussed in Chapter 4.

In this study also, the INDEL markers were applied to the crime and bone samples, in order to evaluate the usefulness of these INDELs in handling problematic samples. The results showed that the INDEL assay returns more genetic information for blood stains and bone samples, compared to STR assay. Even the match probability of the STR is higher, Investigator DIPplex® Kit still gives an acceptable value for forensic identification

e.g. linking different pieces of evidence or re-association of body parts in the case of human identification.

3.6 Conclusion

The chapter shows that the kit is sensitive to relatively low levels of input DNA, compares the effectiveness of two DNA extraction methods in relation to the successful amplification of the kit markers, examines electropherogram features to determine the quality of INDEL profiles, consider the effect of DNA degradation on the production of INDEL profiles and examines the performance of the kit on samples from five Malaysian populations and on casework/bone samples, in comparison to STR analysis.

CHAPTER 4

STATISTICAL ANALYSIS ON POPULATION DATA FROM FIVE MALAYSIAN POPULATIONS

4.1 Overview

In order to calculate the weight of DNA evidence when presenting this in court, knowledge about DNA markers in a relevant population sample is required (Tilmar 2010). Important things that should be studied are: how frequently certain alleles occur in a population, is there any linkage between the markers, can Hardy Weinberg Equilibrium (HWE) be applied when calculating profile frequencies, and if linkage is absent and the markers behave as if the population is in HWE how rare a profile is within different subpopulations. Currently in Malaysia, regardless of the markers, all DNA cases are evaluated based on published allele frequency population databases for the Malays, Malaysian-Chinese (M-Chinese) and Malaysian-Indian (M-Indian). Data for some markers are also available for three major indigenous populations in Sarawak, Iban, Bidayuh and Melanau (Suadi et al., 2007). Kits with 17 STR (including amelogenin) (Othman et al., 2004, Seah et al. 2003) and 16 or 17 YSTR (Chang et al., 2009, Chang et al., 2007) are approved for loading results to the Malaysian Database.

The geneic markers forming a DNA profile are found at different frequencies in each population. The collection of genetic markers is known as a population genetic database. If the frequencies of the genetic markers are high and found in half of the population, the statistical strength or discrimination power of the particular DNA profile is not strong. In contrast, if the genetic markers of the DNA profile are found only in one in a million individuals, then the statistical discrimination power of the DNA profile is very strong. It is therefore, important that the data on the genetic markers of a population genetic database are statistically tested before the data used to determine the occurrence rate of a particular DNA profile in a criminal casework. Neglecting to do so can decrease the strength of the evidence against the defendant.

In this chapter, allelic frequencies were calculated for thirty INDEL markers; following on several statistical parameters of genetic and forensic efficacy parameters were estimated. These includes the heterozygosity, the exact testing for HWE, population

differentiation -F_{ST}, the polymorphism information content (PIC), the power discrimination (DP), the power of exclusion (PE) and the match probability (MP).

The highlight of this study, 30 loci including amelogenin, which have used on other populations (refer to Table 1.3) were also tested on 500 unrelated Malaysian population. Therefore, this is the first study of its kind on the analysis of the allele frequency and genotyping of those INDEL loci in Malaysian population. The data obtained then could be used to initiate Malaysian genetic database for the INDEL polymorphic markers. Moreover, the results of this study can show the utility of those INDEL markers for paternity testing and forensic identifications.

All 30 INDEL markers were reliable in the populations tested (refer to Table 1-3). When testing for potential differential expression across the genotype, each gene considered independently from one another. In other word, ANOVA, HWE or any other test performed on each marker separately. The incidence of false positive (genotypes falsely called differentially expressed when they are not) or false negative (genotypes false called not differentially expressed when they are) are dependent on the number of tests performed and the critical significance level (*p*-value cut off). According to research carried out on Qiagen investigator DIPplex kit, there are no significance deviation from HWE on the markers. No linkage disequilibrium was found between the two loci on the same chromosome, and they are statistically independent. The absence of linkage disequilibrium between all tested loci ensures the stability of the calculations concerning genotype frequencies. The 30 INDEL loci also showed that the INDEL multiplex is suitable for forensic tool (refer to Table 4-13).

4.1.1 Objectives

d. To generate and evaluate of the population genetics data using Qiagen Investigator DIPplex® kit in the five Malaysian sub-populations (approximately 100 samples each). This is because the Malay, M-Chinese, M-Indian, Iban and Bidayuh are the major sub-populations in Malaysia. Moreover, the allele frequency data for these ethnics have been published; 9 STRs (Lim et al. 2001), 15 STRs (Izuan et al. 2005, Seah et al. 2003) and 16 or 17 YSTRs (Chang et al., 2007) from three major ethnic population groups of Malaysia (Malay, Chinese and Indians) and three major native populations in Sarawak: Iban, Bidayuh and Melanau (Suadi et al., 2007, Chang et al., 2009). All forensic DNA cases are evaluated using these DNA population databases.

4.2 Materials and methods

4.2.1 Samples

DNA were extracted from blood stained specimen from the Malays (N = 100), the M-Chinese (N = 100), the M-Indian (N = 100), the Iban (N = 100) and the Bidayuh (N = 100) populations are included in following analysis.

4.2.2 Estimation of population and forensic parameters.

To assess the Investigator DIPplex® Kit's efficiency for application in individual identification, certain forensic parameters were estimated. The forensic parameters polymorphism information content (PIC), random match probability (RMP), power of discrimination (PD) and power of exclusion (PE) were using Powerstats (Tereba 1999). Allele frequencies, expected and observed heterozygosity, as well as probability values (p-values for Hardy-Weinberg equilibrium were estimated using Arlequin (Excoffier, Lischer 2010, Excoffier et al., 2007).

4.2.2.1 The Snipper App

This software was used in this study to demonstrate the performance of the 30 INDEL markers in inferring the ancestral origin of reference samples; Malay, Chinese, Iban and Bidayuh. The classification success was estimated using the 'Verbose cross validation' option by choosing the. Thorough analysis of population data with a custom Excel file method. The INDEL profiles were coded as A/C; where 'A' representing as deletion and 'C' representing as insertion. All profiles were also predicted and classified by choosing the 'Classification and predicted admixture components of multiple profiles with a custom Excel file of populations' method (http://mathgene.usc.es/snipper).

4.3 Results

4.3.1 Population studies

4.3.1.1 Allele frequencies

In this study, a total of 500 individuals comprising 100 each for the Malay, M-Chinese, M-Indian, Iban and Bidayuh were extracted using phenol/chloroform method, quantitated and genotyped using 30 INDEL markers as described in Chapter 2. Table 4-1 showed the allele frequencies obtained for the five population groups. Allele frequencies for the short (deletion; which was coded as **1**) and long (insertion; which was coded as **2**) alleles have been calculated for the Malay, M-Chinese, M-Indian, Iban and Bidayuh, each comprising of 100 individuals. Allele frequencies of the 30 INDEL markers are listed in Table 4-1. It is notable that allele frequencies of most INDEL markers in Malays, M-Chinese, M-Indian, Iban and Bidayuh are very similar to each other. Thus, the minimum allele frequency observed are 0.050 (+HLD122) and 0.055 (+HLD39), in the Iban and Bidayuh respectively, which are highlighted in green. Whilst, the largest allele frequency observed in M-Chinese; 0.920 (-HLD111) and 0.915 (-HLD39); Iban, 0.950 (-HLD122) and Bidayuh, 0.945 (-HLD39) highlighted in yellow.

Table 4-1: Allele frequencies data of Malaysian population. The green highlighted represents the minimum allele frequencies and the yellow represents maximum allele frequencies.

INDEL Locus	Allele Frequencies (-/+ DIP)				
	Malay	M-Chinese	M-Indian	Iban	Bidayuh
HLD77	0.720/0.280	0.600/0.400	0.790/0.210	0.650/0.350	0.705/0.295
HLD45	0.305/0.695	0.420/0.580	0.340/0.660	0.365/0.635	0.295/0.705
HLD131	0.655/0.345	0.625/0.375	0.610/0.390	0.750/0.250	0.695/0.305
HLD70	0.310/0.690	0.370/0.630	0.265/0.735	0.385/0.615	0.295/0.705
HLD6	0.510/0.490	0.575/0.425	0.440/0.560	0.515/0.485	0.565/0.435
HLD111	0.780/0.220	0.920/0.080	0.720/0.280	0.890/0.110	0.875/0.125
HLD58	0.570/0.430	0.540/0.460	0.655/0.345	0.585/0.415	0.750/0.250
HLD56	0.415/0.585	0.480/0.520	0.255/0.745	0.440/0.560	0.415/0.585
HLD118	0.175/0.825	0.060/0.940	0.460/0.540	0.205/0.795	0.100/0.900
HLD92	0.475/0.525	0.570/0.430	0.400/0.600	0.530/0.470	0.490/0.510
HLD93	0.585/0.415	0.405/0.595	0.540/0.460	0.295/0.705	0.360/0.640
HLD99	0.250/0.750	0.120/0.880	0.320/0.680	0.145/0.855	0.215/0.785
HLD88	0.520/0.480	0.545/0.455	0.415/0.585	0.625/0.375	0.690/0.310
HLD101	0.590/0.410	0.565/0.435	0.530/0.470	0.635/0.365	0.580/0.420
HLD67	0.355/0.645	0.220/0.780	0.440/0.560	0.255/0.745	0.300/0.700
HLD83	0.560/0.440	0.585/0.415	0.725/0.275	0.635/0.365	0.565/0.435
HLD48	0.600/0.400	0.635/0.365	0.465/0.535	0.650/0.350	0.555/0.445
HLD114	0.690/0.310	0.735/0.265	0.620/0.380	0.705/0.295	0.725/0.275
HLD124	0.435/0.565	0.485/0.515	0.415/0.585	0.500/0.500	0.510/0.490
HLD122	0.850/0.150	0.800/0.200	0.730/0.270	0.950/0.050	0.825/0.175
HLD125	0.535/0.465	0.600/0.400	0.440/0.560	0.510/0.490	0.490/0.510
HLD64	0.130/0.870	0.110/0.890	0.210/0.790	0.135/0.865	0.135/0.865
HLD81	0.285/0.715	0.160/0.840	0.320/0.680	0.295/0.705	0.260/0.740
HLD136	0.475/0.525	0.495/0.505	0.455/0.545	0.650/0.350	0.615/0.385
HLD133	0.690/0.310	0.640/0.360	0.610/0.390	0.660/0.340	0.735/0.265
HLD97	0.680/0.320	0.630/0.370	0.715/0.285	0.750/0.250	0.720/0.280
HLD40	0.560/0.440	0.305/0.695	0.735/0.265	0.395/0.605	0.320/0.680
HLD128	0.700/0.300	0.760/0.240	0.470/0.530	0.795/0.205	0.825/0.175
HLD39	0.855/0.145	0.915/0.085	0.580/0.420	0.890/0.110	0.945/0.055
HLD84	0.235/0.765	0.215/0.785	0.330/0.670	0.150/0.850	0.110/0.890

Figure 4-1: The graphs of allele frequencies of the Malays, M-Chinese, M-Indian, Iban and Bidayuh population for 30 INDEL loci. The green colour bar represents deletion, while the blue represents insertion.

4.3.1.2 Hardy Weinberg Equilibrium (HWE)

The observed and expected heterozygosity values among the Malay, M-Chinese, M-Indian, Iban and Bidayuh are tabulated in Table 4-2. The minimum observed heterozygosity (highlighted in blue) for Malays, M-Chinese, M-Indian, Iban and Bidayuh are 0.190, 0.100, 0.280, 0.180 and 0.110, respectively and the minimum expected heterozygosity (highlighted in yellow) are 0.227, 0.113, 0.333, 0.095 and 0.104. The mean observed heterozygosity for Malay, M-Chinese, M-Indian, Iban and Bidayuh are 0.413, 0.376, 0.438, 0.394 and 0.379 respectively and the mean expected heterozygosity are 0.427, 0.401, 0.451, 0.397 and 0.389 respectively. A higher heterozygosity means that more allele diversity exists and therefore there will be less chance of a random sample matching.

There were only four INDELs showing departures from HWE ($p<0.05$) which are HLD101 for Malay ($p=0.0009$), HLD133 for M-Indian ($p=0.005$), HLD125 for Iban ($p=0.028$) and HLD93 for Bidayuh ($p=0.014$). However, after Bonferroni correction for multiple testing (Weir, Cockerham 1984, Weir, Cockerham 1984, Weir, Hill 2002, Bland, Altman 1995) the significance level changes from 0.05 to 0.0003. Subsequently, all *p*-values were greater than 0.0003 and all loci including HLD101 for Malay ($p=0.0009$), HLD133 for M-Indian ($p=0.005$), HLD125 for Iban ($p=0.028$) and HLD93 for Bidayuh ($p=0.014$) were therefore in HWE. This established, the 30 loci analysed were statistically independent

Loci	Malay			M-Chinese			M-Indian			Iban			Bidayuh		
	Ho	He	HWE	Ho	He	HWE	Ho	He	HWE	Ho	He	HWE	Ho	He	HWE
HLD77	0.42	0.405	0.805	0.42	0.482	0.215	0.28	0.333	0.13	0.48	0.457	0.663	0.41	0.418	1
HLD45	0.39	0.426	0.478	0.44	0.489	0.411	0.5	0.451	0.373	0.41	0.465	0.278	0.41	0.418	1
HLD131	0.43	0.454	0.659	0.45	0.471	0.675	0.42	0.478	0.293	0.44	0.376	0.11	0.43	0.426	1
HLD70	0.38	0.429	0.253	0.44	0.468	0.667	0.31	0.391	0.042	0.45	0.475	0.673	0.39	0.418	0.629
HLD6	0.5	0.502	1	0.47	0.47	0.686	0.48	0.495	0.839	0.55	0.502	0.424	0.49	0.494	1
HLD111	0.36	0.344	0.775	0.12	0.147	0.109	0.46	0.405	0.216	0.18	0.196	0.327	0.17	0.219	0.041
HLD58	0.54	0.492	0.415	0.54	0.499	0.499	0.49	0.454	0.506	0.49	0.487	1	0.36	0.376	0.79
HLD56	0.45	0.487	0.537	0.42	0.501	0.111	0.35	0.381	0.431	0.48	0.495	0.839	0.51	0.487	0.684
HLD118	0.25	0.29	0.172	0.1	0.113	0.294	0.56	0.499	0.232	0.31	0.327	0.55	0.16	0.18	0.245
HLD92	0.47	0.501	0.553	0.44	0.492	0.309	0.46	0.482	0.678	0.48	0.5	0.691	0.5	0.502	1
HLD93	0.51	0.487	0.683	0.49	0.484	1	0.52	0.499	0.691	0.43	0.418	0.815	0.44	0.463	0.665
HLD99	0.32	0.376	0.178	0.22	0.212	1	0.52	0.437	0.068	0.29	0.249	0.213	0.25	0.339	0.014
HLD88	0.42	0.501	0.112	0.49	0.498	1	0.45	0.487	0.536	0.51	0.471	0.522	0.36	0.429	0.108
HLD101	0.32	0.486	0.0009	0.43	0.494	0.223	0.48	0.5	0.694	0.39	0.465	0.13	0.56	0.489	0.156
HLD67	0.43	0.46	0.52	0.34	0.344	1	0.52	0.495	0.685	0.41	0.381	0.599	0.44	0.422	0.812
HLD83	0.48	0.495	0.839	0.49	0.487	1	0.41	0.4	1	0.39	0.465	0.129	0.59	0.494	0.068
HLD48	0.42	0.482	0.214	0.39	0.465	0.13	0.51	0.5	1	0.44	0.457	0.826	0.51	0.496	0.84
HLD114	0.44	0.429	1	0.41	0.391	0.797	0.42	0.473	0.29	0.41	0.418	1	0.37	0.4	0.458
HLD124	0.51	0.494	0.839	0.51	0.502	1	0.43	0.487	0.302	0.46	0.502	0.426	0.52	0.502	0.84
HLD122	0.22	0.256	0.224	0.26	0.321	0.063	0.42	0.396	0.619	0.08	0.095	0.21	0.31	0.29	0.729
HLD125	0.53	0.5	0.684	0.46	0.482	0.679	0.42	0.495	0.156	0.62	0.502	0.028	0.42	0.502	0.112
HLD64	0.22	0.227	0.665	0.16	0.196	0.089	0.38	0.333	0.228	0.21	0.234	0.376	0.27	0.234	0.204
HLD81	0.41	0.409	1	0.22	0.27	0.122	0.36	0.437	0.105	0.39	0.418	0.63	0.38	0.386	1
HLD136	0.51	0.501	1	0.47	0.502	0.553	0.47	0.498	0.686	0.38	0.457	0.122	0.43	0.475	0.398
HLD133	0.46	0.429	0.639	0.48	0.463	0.829	0.46	0.478	0.833	0.46	0.451	1	0.37	0.391	0.612
HLD97	0.46	0.437	0.651	0.42	0.468	0.389	0.29	0.409	0.005	0.34	0.376	0.423	0.34	0.405	0.135
HLD40	0.56	0.495	0.222	0.41	0.426	0.813	0.43	0.391	0.44	0.53	0.48	0.402	0.46	0.437	0.65
HLD128	0.42	0.422	1	0.34	0.366	0.582	0.46	0.5	0.427	0.35	0.327	0.756	0.23	0.29	0.072
HLD39	0.19	0.249	0.03	0.17	0.156	1	0.46	0.489	0.681	0.22	0.196	0.599	0.11	0.104	1
HLD84	0.37	0.361	1	0.29	0.339	0.149	0.44	0.444	1	0.24	0.256	0.452	0.18	0.196	0.327
Mean	**0.413**	**0.427**		**0.376**	**0.401**		**0.438**	**0.451**		**0.394**	**0.397**		**0.379**	**0.389**	

Table 4-2 Observed (Ho) and expected (He) heterozygosity and p values form exact test for HWE across 30 INDEL markers for Malaysian population. Blue represents minimum Ho; yellow represents minimum He; and green $p < 0.05$.

4.3.1.3 Exact test for Linkage Disequilibrium

The pairwise linkage disequilibrium (LD) describes the degree of association between particular alleles at different genetic loci. Linkage disequilibrium can result when polymorphic genetic loci are situated close to each other on the same chromosome. In this research, the LD between the loci for each population group was also tested as shown in Table 4-3. 35 pairs in Malays, 38 pairs in M-Chinese, 19 pairs in M-Indian, 32 pairs in Iban and 31 pairs in Bidayuh were detected demonstrating significant LD $p<0.05$, which are highlighted with yellow boxes. After applying a Bonferroni correction at $p<0.0001149$, the data revealed that overall, there were no significant linkage disequilibrium among the INDEL loci on the same chromosomes, and consequently, the assumption of independence among all the 30 markers was reasonable for the Malay, M-Chinese, M-Indian, Iban and Bidayuh groups. The significant LD of all pairs of loci for the Malay, M-Chinese, M-Indian, Iban and Bidayuh were shown in Figure 4-2.

Table 4-3: The table below shows significant linkage disequilibrium at p<0.05 in Malays, M-Chinese, M-Indian, Iban and Bidayuh.

Marker Code	Loci INDEL	Linked loci per locus				
		Malay	M-Chinese	M-Indian	Iban	Bidayuh
0	HLD77	2	3	2	0	3
1	HLD45	0	4	1	2	3
2	HLD131	4	2	3	5	0
3	HLD70	4	3	1	2	0
4	HLD6	2	2	1	1	2
5	HLD111	4	1	0	3	3
6	HLD58	3	4	1	3	1
7	HLD56	1	6	0	1	3
8	HLD118	0	3	1	3	2
9	HLD92	3	3	1	3	3
10	HLD93	1	3	1	3	1
11	HLD99	3	1	2	1	2
12	HLD88	3	0	3	2	1
13	HLD101	0	1	1	2	2
14	HLD67	4	4	1	4	4
15	HLD83	5	4	2	0	1
16	HLD48	1	3	2	1	3
17	HLD114	1	1	2	1	2
18	HLD124	3	4	1	3	2
19	HLD122	1	2	0	4	2
20	HLD125	1	2	1	1	4
21	HLD64	4	3	0	3	0
22	HLD81	4	4	2	2	1
23	HLD136	3	3	1	1	4
24	HLD133	3	1	1	0	0
25	HLD97	0	1	1	2	3
26	HLD40	3	2	3	2	3
27	HLD128	3	1	0	5	3
28	HLD39	3	4	2	2	3
29	HLD84	1	1	1	2	1
Total linked loci		70 (35 pairs)	76 (38 pairs)	38 (19 pairs)	64 (32 pairs)	62 (31 pairs)

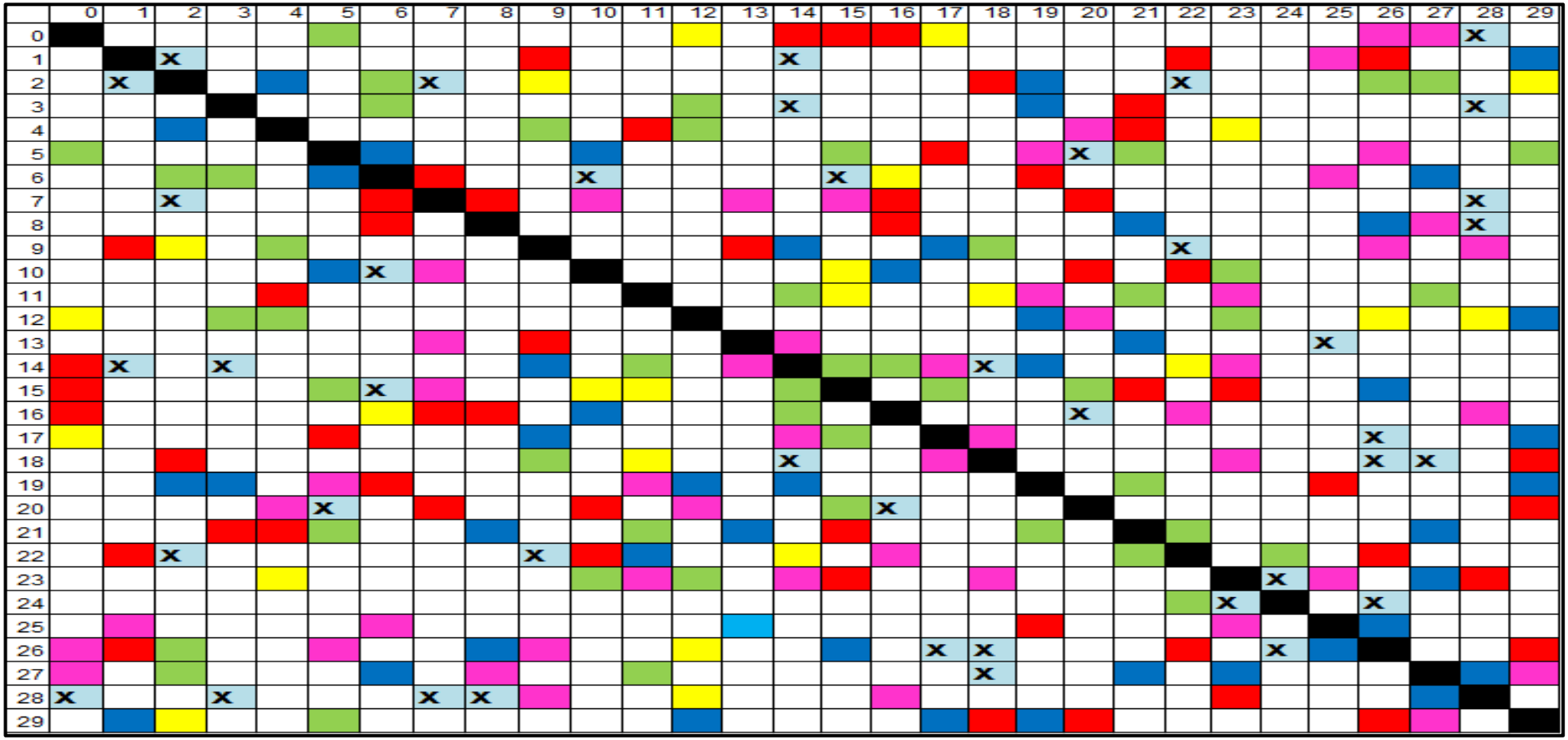


Figure 4-2: Significant LD detected in the five Malaysian ethnic groups. All coloured boxes represents significant LD with $p < 0.05$, but not significant after Bonferroni correction ($p < 0.0001149$).

4.3.1.4 Population differentiation test

Population differentiation is calculated from allele frequencies and to determine whether genetic differentiation exists between populations (Raymond, Rousset 1995). When the null hypothesis is no differentiation amongst populations, where $p < 0.05$, it is said to show significant differentiation.

Study shows the Malaysian populations; Malays (**M**), M-Chinese (**C**), M-Indian (**Ind**), Iban (**Ib**) and Bidayuh (**By**), are compared within each other on the 30 INDEL loci. The results for each population per locus is summarized in Table 4-4. Statistically significant differences (after corrections Bonferroni's correction; $p < 0.01$) were observed.

Table 4-4 shows significant differences in the M-C population pairs at 8 loci; M-In at 9 loci; M-Ib at 6 loci; M-By at 9 loci; C-Ind at 17 loci; C-Ib at 4 loci; C vs By at 1 locus; Ind-Ib at 17 loci, Ind-By at 12 loci and Ib-By at 3 loci. The exact population differentiation test indicates differences between the five Malaysian populations. The C-Ind and Ind-Ib pairs showed most differentiated populations, followed by Ind-By. The M-C, M-Ind, M-Ib and M-By displayed similar significant differences across all loci. However, the C-By pair showed no significant differentiation for all loci except for HLD58. The C-Ib and Ib-By pairs respectively displayed significant differentiation at 4 (HLD118, HLD122, HLD81 and HLD136) and 3 loci (HLD58, 1 HLD18 and HLD122).

Among the 30 loci, the highest ethnic diversity was obtained at 3 loci (HLD45, HLD118 and HLD40) with significant differences found within the Malaysian population, followed by HLD93 at 6 population pairs and HLD111, HLD122, HLD136, HLD128 and HLD39 (5 population pairs). The lowest ethnic diversity found at 7 loci (HLD70, HLD101, HLD48, HLD124, HLD64, HLD133 and HLD97), where no significant differentiation within the Malaysian population pairs studied. The results showed that, ethnic diversity varied between different INDEL markers. Some loci showed no significant difference at all, while others would have significant differences between pairs with relatively close relationship.

Table 4-4: Population differentiation within the Malaysian populations; Malay (M), M-Chinese (C), M-Indian (Ind), Iban (Ib) and Bidayuh (By). Figure in bold, italic and highlighted in pink indicate a *p*-value less than 0.05, figures in bold italic and highlighted in green indicate a significant *p*-value after Bonferroni correction for 5 populations (i.e. 0.01). The exact test was carried out with 300,000 Markov steps. [- - -] indicates no significant difference.

Indel	M vs C	M vs Ind	M vs Ib	M vs By	C vs Ind	C vs Ib	C vs By	Ind vs Ib	Ind vs By	Ib vs By
HLD77	- - -	- - -	- - -	- - -	***0.0000***	- - -	- - -	***0.0020***	- - -	- - -
HLD45	***0.0000***	***0.0003***	***0.0020***	***0.0000***	***0.0000***	- - -	- - -	***0.0000***	***0.0000***	- - -
HLD131	- - -	- - -	- - -	- - -	- - -	- - -	- - -	***0.0029***	- - -	- - -
HLD70	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -
HLD6	- - -	- - -	- - -	- - -	***0.0094***	- - -	- - -	- - -	- - -	- - -
HLD111	***0.0001***	- - -	***0.0041***	- - -	***0.0000***	- - -	- - -	***0.0000***	***0.0004***	
HLD58	- - -	- - -	- - -	***0.0001***	- - -	- - -	***0.0000***	- - -	- - -	***0.0001***
HLD56	- - -	***0.0007***	- - -	- - -	***0.0000***	- - -	- - -	***0.0005***	***0.0013***	- - -
HLD118	***0.0007***	***0.0000***	- - -	- - -	***0.0000***	***0.0000***	- - -	***0.0000***	***0.0000***	***0.0041***
HLD92	- - -	- - -	- - -	- - -	***0.0013***	- - -	- - -	- - -	- - -	- - -
HLD93	***0.0003***	- - -	***0.0000***	***0.0000***	***0.0068***	- - -	- - -	***0.0000***	***0.0003***	- - -
HLD99	***0.0013***	- - -	- - -	- - -	***0.0000***	- - -	- - -	***0.0001***	- - -	- - -
HLD88	- - -	- - -	- - -	***0.0012***	- - -	- - -	- - -	***0.0000***	***0.0000***	- - -
HLD101	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -
HLD67	***0.0045***	- - -	- - -	- - -	***0.0000***	- - -	- - -	***0.0001***	- - -	- - -
HLD83	- - -	***0.0015***	- - -	- - -	***0.0035***	- - -	- - -	- - -	***0.0015***	- - -
HLD114	- - -	***0.0092***	- - -	- - -	***0.0015***	- - -	- - -	***0.0001***	- - -	- - -
HLD48	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -
HLD124	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -
HLD122	- - -	***0.0053***	***0.0011***	- - -	- - -	***0.0000***	- - -	***0.0000***	- - -	***0.0002***
HLD125	- - -	- - -	- - -	- - -	***0.0013***	- - -	- - -	- - -	- - -	- - -
HLD64	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -
HLD81	***0.0038***	- - -	- - -	- - -	***0.0003***	***0.0016***	- - -	- - -	- - -	- - -
HLD136	- - -	- - -	***0.0001***	***0.0065***	- - -	***0.0022***	- - -	***0.0000***	***0.0016***	- - -
HLD133	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -
HLD97	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -
HLD40	***0.0000***	***0.0002***	***0.0013***	***0.0000***	***0.0000***	- - -	- - -	***0.0000***	***0.0000***	- - -
HLD128	- - -	***0.0001***	- - -	***0.0048***	***0.0000***	- - -	- - -	***0.0000***	***0.0000***	- - -
HLD39	- - -	***0.0000***	- - -	***0.0034***	***0.0000***	- - -	- - -	***0.0000***	***0.0000***	- - -
HLD84	- - -	- - -	- - -	***0.0016***	- - -	- - -	- - -	***0.0000***	***0.0000***	- - -

4.3.1.5 Genetic differentiation-F_{st}

Table 4-5 shows the genetic differentiation as calculated between the five sub populations using Arlequin software. The genetic distance estimated results were similar to the results show in section 4.3.1.6. The genetic distance is to measure of the difference in the allele frequency between two populations. The greatest genetic distance was between M-Chinese and M-Indian (0.07003). The lowest F_{st} was between Iban and Bidayuh (0.00608); which indicated these populations were closer to each other. The greater the genetic distance between populations, the less breeding there is between them and the more isolated they are from one another. Most F_{st} values between the Malay and M-Chinese, M-Indian, Iban and Bidayuh were significantly low, indicated that there are more breeding between them and are less isolated.

Table 4-5 Population comparison estimated between the Malays, M-Chinese, M-Indian, Iban and Bidayuh using genetic differentiation F_{st} (Weir, Cockerham 1984, Weir, Hill 2002).

	Malay	M-Chinese	M-Indian	Iban	Bidayuh
Malay	-				
M-Chinese	0.01560	-			
M-Indian	0.02652	0.07003	-		
Iban	0.01597	0.01024	0.06152	-	
Bidayuh	0.01662	0.01280	0.06681	0.00608	-

4.3.1.6 Analysis of Molecular Variance (AMOVA)

In this research, Wright's F-statistics was analysed to measure the population substructure effects (Weir, Cockerham 1984) in order to quantify the strength of matching DNA profiles (Weir, Cockerham 1984, Weir, Cockerham 1984, Weir, Hill 2002, Weir, Hill 2002). In this research, the Malaysian population was divided into three subgroups (Group 1- Malay, Iban and Bidayuh; Group 2- M-Chinese and Group 3- M-Indian). AMOVA results revealed that most of the molecular variation was due to variation within populations (96%) rather than among populations within groups (1.2%), with all fixation indexes (F_{ST}) in range 0.00677 to 0.19815. This indicated that random mating is occurring (moderate) but no genetic divergence within the population.

F_{IS} (Individual within the Subpopulation) is also observed in this research in order to measure the extent of genetic inbreeding within subpopulations. It is also known as inbreeding coefficient. The value ranges between 0 and +1. In this study, the F_{IS} value for the Malaysian population presented in the Table 4-6 below. The Iban has the lowest F_{IS} value 0.007. As compared to other sub populations, the Chinese showed higher inbreeding coefficient. Most of the F_{IS} values are near to 0; maybe due to consanguineous marriages. Overall, the average F_{IS} value for Malaysian population has been calculated as 0.032 and significant value is 0.00010.

Table 4-6**:** The AMOVA design and results from Malaysian populations which divided by three subgroups ([1] Malay, Iban, Bidayuh; [2] Chinese and [3] M-Indian).

Population	Name	F_{IS}	P value
1	Malay	0.035	0.107
2	Iban	0.007	0.390
3	Bidayuh	0.027	0.161
4	Chinese	0.062	0.011
5	M-M-Indian	0.027	0.162

Table 4-7: The genetics coefficients (F_{ST}) and inbreeding coefficient (FIS) values for the subpopulations of Malaysian. The negative values (highlighted in pink) indicated outbreeding occurred. The Average F-statistics over all loci; F_{ST}: 0.03915; F_{IS}: 0.03210; F_{IT}: 0.06984

Loci	AMOVA results for polymorphic loci		Population specific FIS indices per polymorphic locus (absolute values)					
	F_{ST}	P Value	Average F_{IS}	Malay	M-Chinese	M-Indian	Iban	Bidayuh
1	0.0269	0.0009	0.0414	-0.0367	0.1299	0.1610	-0.0500	0.0193
2	0.0083	0.0451	0.0450	0.0851	0.1019	-0.1091	0.1205	0.0193
3	0.0112	0.0232	0.0166	0.0536	0.0450	0.1222	-0.1684	-0.0093
4	0.0073	0.0404	0.0984	0.1167	0.0612	0.2090	0.0547	0.0674
5	0.0099	0.0542	-0.0021	0.0046	0.0434	0.0310	-0.0960	0.0082
6	0.0563	0.0000	0.0189	-0.0439	0.1896	-0.1359	0.0857	0.2276
7	0.0198	0.0000	-0.0473	-0.0966	-0.0820	-0.0792	-0.0041	0.0450
8	0.0378	0.0002	0.0618	0.0782	0.1636	0.0838	0.0310	-0.0453
9	0.1895	0.0000	0.0223	0.1391	0.1184	-0.1223	0.0540	0.1161
10	0.0168	000159	0.0524	0.0627	0.1074	0.0467	0.0416	0.0046
11	0.0415	0.0000	-0.0159	-0.0453	-0.0117	-0.0417	-0.0288	0.0502
12	0.0431	0.0000	0.0093	0.1516	-0.0367	-0.1900	-0.1647	0.2641
13	0.0454	0.0000	0.0670	0.1636	0.0170	0.0782	-0.0830	0.1633
14	0.0021	0.3713	0.1057	0.3430	0.1302	0.0416	0.1636	-0.1445
15	0.0393	0.0000	-0.0170	0.0660	0.0143	-0.0502	-0.0741	-0.0426
16	0.0202	0.0031	-0.0069	0.0310	-0.0041	-0.0232	0.1636	-0.1955
17	0.0234	0.0027	0.0552	0.1300	0.1636	-0.0200	0.0380	-0.0274
18	0.0083	0.1075	0.0303	-0.0235	-0.0475	0.1136	0.0193	0.7710
19	0.0021	0.2483	0.0238	-0.0325	-0.0159	0.1194	0.0850	-0.0354
20	0.0465	0.0000	0.0515	0.1422	0.1923	-0.0604	0.1280	-0.0686
21	0.0139	0.0287	0.0131	-0.0602	0.0467	0.1526	-0.2358	0.1646
22	0.0119	0.0497	-0.0106	0.0324	0.1878	-0.1404	0.1058	-0.1512
23	0.0223	0.0053	0.0846	-0.0010	0.1864	0.1777	0.0674	0.0175
24	0.0250	0.0000	0.0724	-0.0175	0.0649	0.0573	0.1697	0.0970
25	0.0070	0.0783	-0.0074	-0.0703	-0.0367	0.0382	-0.0200	0.0552
26	0.0067	0.1326	0.1186	-0.0520	0.1041	0.2930	0.0983	0.1616
27	0.1523	0.0000	-0.0719	-0.1314	0.0379	-0.0989	-0.1039	-0.0520
28	0.1201	0.0000	0.0565	0.0050	0.0730	0.0817	-0.0688	0.2083
29	0.1982	0.0000	0.0390	0.2385	-0.0879	0.0608	-0.1186	-0.0532
30	0.0459	0.0000	0.0491	-0.0240	0.1458	0.1000	0.0638	0.0857

4.3.1.7 Population assignment of individual genotypes.

The efficiency of the 30 INDEL markers for assigning individuals to five population groups were further evaluated using Snipper app suite v2.0. As presented in Table 4-8, 86% of the samples with Bidayuh origin were classified as Bidayuh by the algorithm, with only about 1% of the samples having been misclassified as Malays and Iban respectively and 12% for the M-Chinese. 65% success rates were obtained for the M-Indian samples with only 3% of these samples having been misclassified as belonging to the Malays, Bidayuh (14%) and Iban (20%).

Within Malay populations, only 13% of the genotypes were correctly classified, whereas 27% and 29% were correctly classified within M-Chinese and Iban origin groups. Similar proportions of Malay were misclassified in M-Chinese (24%) and M-Indian (28%), then Iban (17%) and Bidayuh (18%). However, 33% and 31% were misclassified as Iban and Bidayuh within M-Chinese origin compare to M-Indian (9%). 66% which have misclassified as Bidayuh within Iban but only small proportions were misclassified as Malay (3%) and M-Chinese (2%). In general, the populations which were carried out in this study are closely related to be properly differentiated with small sets of 30 autosomal INDEL markers. The 46-plex INDEL set can be used in the future as there are designed for broad-scale continental comparisons. Cautions should be carried out as a forensic ancestry marker sets must be small to ensure sensitivity and stability to work with difficult DNA but for this reason there are not suited to study geographically origin. Cardoso et al., 2017 reported, even though this INDELs set efficiently differentiate between main ancestries, it still does not allow an accurate separation at a local level and, for the time being, their combination with other informative markers is needed to maximize the power to accurately differentiate populations with close genetic ancestry.

Table 4-8: Estimated classification success for all tested individuals during cross validation with Snipper app suite v2.0.

	Malay	M-Chinese	M-Indian	Iban	Bidayuh
Malay origin	13%	24 %	28 %	17 %	18 %
M-Chinese origin	0 %	27 %	9 %	33 %	31 %
M-Indian origin	1 %	0 %	65 %	20 %	14 %
Iban origin	3 %	2 %	0 %	29 %	66 %
Bidayuh origin	1 %	12 %	0 %	1 %	86 %

The classification probabilities obtained are also represented as a pentagonal plot in Figure 4-3. Most of the individuals from the population can be seen clustered together. The PCA (principle component analysis) of individuals from the Iban and Bidayuh are almost fully overlapping clusters, an indication of genetic similarities between the two populations. The Malays and M-Chinese are dispersed throughout the plot, signifying genetic similarities shared with other populations whereas the M-Indian are the most divergent population and also offers the best classification success.

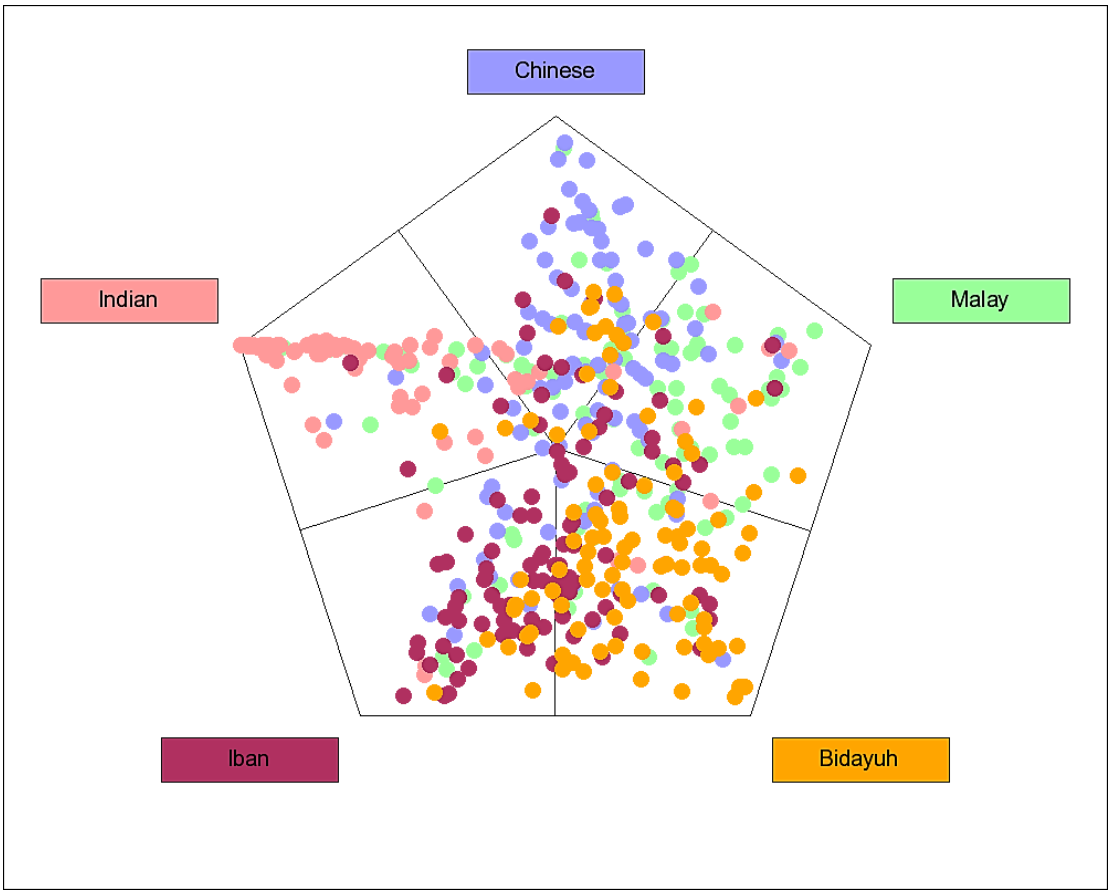


Figure 4-3: A pentagonal plot classifying the individuals.

In conjunction with that, another evaluation of the Malaysian population was carried out but individuals of Iban and Bidayuh (as the F_{ST} for both are low) were excluded from the training set for increasing the differences. Table 4-8 shows that the success ratio (values in bold) for the M-Indian is far higher at 98% than in the Chinese (76%) and the Malays (20%). Indirectly, this also indicates these INDEL markers have high probabilities in discriminating M-Indian from the Malays and the M-Chinese with only a low probability of differentiating between the M-Chinese and Malays. Triangular plot

(Figure 4-4) shows only several Malay and M-Chinese individuals are misclassified as M-Indian and vice versa.

Table 4-9: Estimation with cross validation of the Malaysian population without Iban and Bidayuh using Snipper app suite v2.0.

	Malay	Chinese	M-M-Indian
Malay origin	20%	45%	35%
Chinese origin	0%	76%	24%
M-M-Indian origin	2%	0%	98%

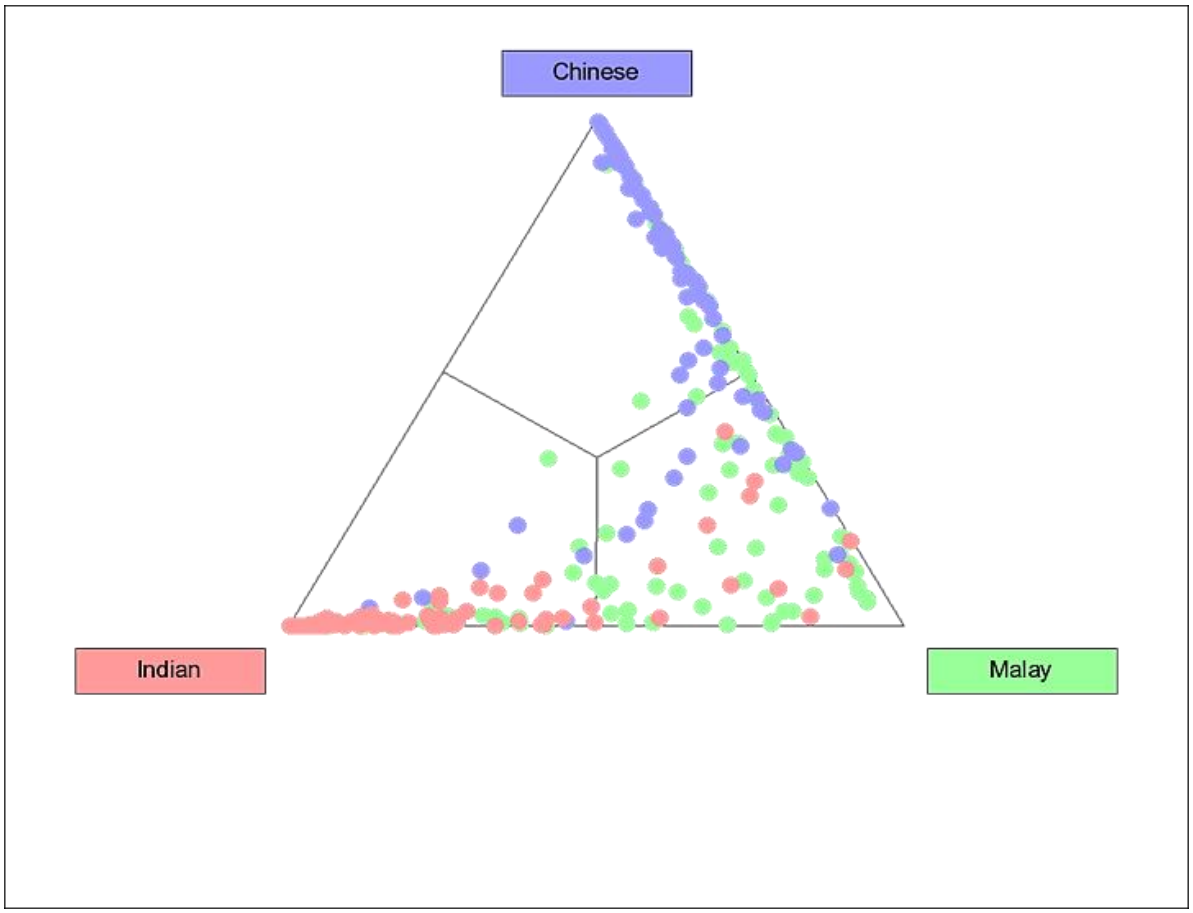


Figure 4-4: A triangular plot classifying of the Malaysian population excluded the Iban and Bidayuh individuals.

4.3.2 Forensic Parameters

Table 4-10 lists the estimated values for random match probabilities (RMD), power of discrimination (PD), power of exclusion (PE) and polymorphic (PIC), for Malays, M-Chinese, M-Indian, Iban and Bidayuh populations.

PIC is a measure of the polymorphic level of a locus (Goodwin, Linacre & Hadi 2011). It was reported that a diallelic locus can have maximum PIC value of 0.375 (Hildebrand et al. 1994). Most of the loci was highly polymorphic (PIC are greater than 0.25) in the five populations. In Malays, there are three markers which were found less informative (HLD122, HLD64 and HLD84), in M-Chinese five markers (HLD111, HLD118, HLD99, HLD81 and HLD39); in Iban six markers (HLD111, HLD99, HLD122, HLD64, HLD39 and HLD84) and lastly in Bidayuh, five markers (HLD111, HLD118, HLD64, HLD39 and HLD84). These are bold and highlighted in pink. In the M-Indian population, all of the loci were revealed to be highly polymorphic, values between 0.28 and 0.37.

The highest power of discrimination (PD) was found on loci HLD88 for Malays, HLD56 for M-Chinese and HLD125 for Bidayuh (all PD = 0.655), all were bold and highlighted with blue. The combined power of discrimination for the DIPplex kit for the Malaysian population were all >99.9999 %, allowing an acceptable level of discrimination in forensic cases. These values indicate a low likelihood of random match confirming that the Investigator DIPplex kit is an effective tool for identification of individuals (Akhteruzzaman et al. 2013, Akhteruzzaman et al. 2013) in the Malaysian population. The assessed combined probability of exclusions (CPE) for the Malaysian population groups were all 99.99 % (Table 4-10).

The combined match probability (CPM) of all 30 loci were 1 in 5.25×10^{-12}, 1 in 2.70×10^{-11}, 1 in 1.79×10^{-12}, 1 in 4.87×10^{-11}, 1 in 7.6×10^{-11} for Malays, M-Chinese, M-Indian, Iban and Bidayuh, respectively. The values are higher than the 15 STRs and SNPs typing on the Malaysian populations (see Table 4-11).

Table 4-10: Forensic parameters calculated for INDELs for each Malaysian ethnic group. PM represents match probability, PD represents power of discrimination and PIC represents polymorphism information content. The bold and highlighted in blue shows the highest power of discrimination while bold and highlighted in pink shows loci which are less informative.

Loci	Malay				Chinese				M-M-Indian				Iban				Bidayuh			
	PM	PD	PE	PIC	PM	PD	PE	PIC	PM	PD	PE	PIC	PM	PD	PE	PIC	PM	PD	PE	PI
HLD77	0.441	0.559	0.127	0.32	0.365	0.635	0.127	0.36	0.506	0.494	0.056	0.28	0.411	0.589	0.171	0.35	0.426	0.574	0.120	0.
HLD45	0.414	0.586	0.108	0.33	0.363	0.637	0.140	0.37	0.426	0.574	0.188	0.35	0.379	0.621	0.120	0.36	0.426	0.574	0.120	0.
HLD131	0.395	0.605	0.133	0.35	0.385	0.615	0.147	0.36	0.369	0.631	0.127	0.36	0.475	0.525	0.140	0.30	0.423	0.577	0.133	0.
HLD70	0.409	0.591	0.102	0.34	0.384	0.616	0.140	0.36	0.444	0.555	0.068	0.31	0.380	0.620	0.147	0.36	0.422	0.578	0.108	0.
HLD6	0.375	0.625	0.188	0.37	0.373	0.627	0.163	0.37	0.373	0.627	0.171	0.37	0.404	0.600	0.235	0.37	0.379	0.621	0.179	0.
HLD111	0.491	0.509	0.091	0.28	0.754	0.246	0.012	**0.14**	0.454	0.546	0.155	0.32	0.673	0.327	0.025	**0.18**	0.655	0.345	0.022	**0.**
HLD58	0.407	0.593	0.225	0.37	0.401	0.599	0.225	0.37	0.418	0.582	0.179	0.35	0.385	0.615	0.179	0.37	0.459	0.541	0.091	0.
HLD56	0.368	0.632	0.147	0.37	0.345	**0.655**	0.127	0.37	0.454	0.546	0.086	0.31	0.373	0.627	0.176	0.37	0.395	0.605	0.196	0.
HLD118	0.555	0.445	0.045	0.25	0.802	0.198	0.008	**0.11**	0.415	0.586	0.246	0.37	0.508	0.492	0.068	0.27	0.698	0.302	0.020	**0.**
HLD92	0.363	0.637	0.163	0.37	0.360	0.640	0.140	0.37	0.377	0.623	0.155	0.36	0.367	0.633	0.171	0.37	0.375	0.625	0.188	0.
HLD93	0.395	0.605	0.196	0.37	0.388	0.612	0.179	0.37	0.389	0.611	0.206	0.37	0.431	0.569	0.133	0.33	0.390	0.610	0.140	0.
HLD99	0.459	0.541	0.072	0.30	0.641	0.359	0.035	**0.19**	0.450	0.550	0.206	0.34	0.588	0.412	0.060	**0.22**	0.506	0.494	0.045	0.
HLD88	0.345	**0.655**	0.127	0.37	0.374	0.626	0.179	0.37	0.368	0.632	0.147	0.37	0.411	0.589	0.196	0.36	0.407	0.593	0.091	0.
HLD101	0.350	0.650	0.072	0.37	0.356	0.644	0.133	0.37	0.367	0.633	0.171	0.37	0.375	0.625	0.108	0.36	0.423	0.577	0.246	0.
HLD67	0.389	0.610	0.133	0.35	0.490	0.510	0.081	0.28	0.393	0.607	0.206	0.37	0.462	0.538	0.120	0.31	0.430	0.570	0.140	0.
HLD83	0.373	0.627	0.171	0.37	0.385	0.615	0.179	0.37	0.443	0.557	0.120	0.32	0.375	0.625	0.108	0.36	0.441	0.559	0.279	0.
HLD48	0.365	0.635	0.127	0.36	0.375	0.625	0.108	0.36	0.383	0.617	0.196	0.37	0.395	0.605	0.140	0.35	0.386	0.614	0.196	0.
HLD114	0.423	0.577	0.140	0.34	0.453	0.547	0.120	0.31	0.373	0.627	0.127	0.36	0.426	0.574	0.120	0.33	0.437	0.563	0.097	0.
HLD124	0.389	0.611	0.196	0.37	0.381	0.619	0.196	0.37	0.362	0.638	0.133	0.37	0.357	0.643	0.155	0.38	0.386	0.614	0.206	0.
HLD122	0.598	0.402	0.035	**0.22**	0.521	0.479	0.048	0.27	0.450	0.550	0.126	0.32	0.835	0.165	0.006	**0.09**	0.545	0.455	0.068	0.
HLD125	0.394	0.606	0.215	0.37	0.377	0.623	0.154	0.36	0.352	0.648	0.127	0.37	0.457	0.543	0.316	0.37	0.345	**0.655**	0.127	0.
HLD64	0.626	0.374	0.035	**0.20**	0.683	0.317	0.020	**0.18**	0.505	0.495	0.102	0.28	0.623	0.377	0.033	**0.21**	0.606	0.394	0.052	**0.**
HLD81	0.435	0.565	0.120	0.32	0.584	0.416	0.035	0.23	0.399	0.601	0.091	0.34	0.422	0.578	0.108	0.33	0.452	0.548	0.102	0.
HLD136	0.381	0.619	0.196	0.37	0.361	0.639	0.163	0.37	0.365	0.635	0.166	0.37	0.383	0.618	0.102	0.35	0.374	0.626	0.133	0.
HLD133	0.430	0.570	0.155	0.34	0.405	0.595	0.171	0.35	0.382	0.618	0.154	0.36	0.409	0.591	0.155	0.35	0.446	0.554	0.097	0.
HLD97	0.422	0.578	0.155	0.34	0.378	0.622	0.127	0.36	0.429	0.571	0.060	0.32	0.458	0.541	0.081	0.30	0.430	0.570	0.081	0.
HLD40	0.418	0.582	0.246	0.37	0.418	0.582	0.120	0.33	0.458	0.542	0.133	0.31	0.413	0.587	0.215	0.36	0.422	0.578	0.155	0.
HLD128	0.425	0.575	0.127	0.33	0.469	0.531	0.081	0.30	0.359	0.641	0.155	0.37	0.508	0.492	0.086	0.27	0.567	0.439	0.038	0.
HLD39	0.616	0.384	0.027	**0.22**	0.718	0.282	0.022	**0.14**	0.370	0.630	0.155	0.37	0.657	0.343	0.035	**0.18**	0.804	0.196	0.010	**0.**
HLD84	0.476	0.524	0.097	0.29	0.499	0.501	0.060	0.28	0.408	0.592	0.140	0.34	0.591	0.409	0.041	**0.22**	0.673	0.327	0.025	**0.**
Combined calculation	MP = 1 in 7.20 x 10^{-12} PD > 99.9999 % Ave. PE = 99.99 %				MP = 1 in 2.70 x 10^{-11} PD > 99.9999 % Ave. PE = 99.99 %				MP = 1 in 1.79 x 10^{-12} PD > 99.9999 % Ave. PE = 99.99 %				MP = 1 in 4.87 x 10^{-11} PD > 99.9999 % Ave. PE = 99.99 %				MP = 1 in 7.6 x 10^{-11} PD > 99.9999 % Ave. PE = 99.99 %			

INDELs data obtained in this study were compared to the existing Malaysia STR and SNPs allele frequency database at Department of Chemistry Malaysia. The results show that the 30 INDELs data gave a lower matching probability compared to the STRs and 52 SNPs. but still has potential value when analysing highly degraded material. The data herein support that the Investigator DIPplex® kit provides a powerful supplement and sufficient to use as stand-alone kit for human identity testing. However, the rest of the parameters remain similar (shown in Table 4-11).

Table 4-11: Forensic parameters calculated for 30 INDELs from each Malaysian ethnic group. CMP represents combined match probability, CPD represents combined power of discrimination and CPE represents combined probability of exclusion for different kit used on the Malaysian population.

Forensic multiplex	Loci	Combined match probability (CMP)	Combined power of discrimination (CPD)	Combined probability of exclusion (CPE)	Ethnics in Malaysian Population
Investigator DIPplex kit (in this study)	30	1 in 7.20 x 10^{-12} 1 in 2.70 x 10^{-11} 1 in 1.79 x 10^{-12} 1 in 4.87 x 10^{-11} 1 in 7.60 x 10^{-11}	99.9999 % 99.9999 % 99.9999 % 99.9999 % 99.9999 %	99.99 % 99.99 % 99.99 % 99.99 % 99.99 %	Malay M-Chinese M-Indian Iban Bidayuh
AmpFISTR Profiler[1]	7	N/A	All > 99.90 %	99.99 %	Malay, M-Chinese and M-Indian
AmpFISTR Identifiler[2]	15	2.6 x 10^{-17} 7.0 x 10^{-16} 3.6 x 10^{-17}	99.9999 % 99.9999 % 99.9999 %	99.99 % 99.99 % 99.99 %	Malay M-Chinese M-Indian
AmpFISTR Identifiler[3]	15	>0.9999	0.9999	99.99 %	Malay
AmpFISTR Identifiler[4]	15	3.29 x 10^{-17} 6.55 x 10^{-17} 1.52 x 10^{-16}	99.9999 % 99.9999 % 99.9999 %	99.99 % 99.99 % 99.99 %	Iban Bidayuh Melanau
SNP[5]	52	4.10 x 10^{-19} 5.40 x 10^{-18} 1.74 x 10^{-19}	>99.999%	99.99 % 99.99 % 99.99 %	Malay M-Chinese M-Indian

1Lim et al., 2001; 2 Seah et al., 2003; 3 Maruyama et al., 2008; 4 Suadi et al., 2007; 4 Sharizah Alimat PhD thesis 2014. N/A represents non-applicable

4.3 Discussion

The vast majority of DNA molecules (over 99.7%) are the same between people. It is the small fraction of DNA (0.3% or about 10 million nucleotides) that differ between people and make individuals unique (Krishnamurthy, Manoj & Pagare 2011). The amount of variation, however, is not constant throughout the human genome. Regions containing INDELs are of interest to the forensic scientist because they are abundance, low mutation rate, have very short fragments and easy to interpret which make them effective for human identification purposes. There are however, similarities if we compare the DNA material of two individuals. An individual's DNA profile becomes rarer as more loci are typed. The frequency of a given profile is estimated from the frequencies of the allele of the loci investigated with the aim of answering the following question: "What is the probability of the individual having the same DNA profile in the Malaysian population?"

4.3.1 Distribution of allele frequencies

The allelic frequencies among the Malaysian population at the 30 INDEL markers are illustrated in Table 4-1. The frequencies observed in the Malays, M-Chinese, M-Indian, Iban and Bidayuh subpopulations are similar, ranging from a minimum of 0.050 and maximum 0.950. So far, in comparison of INDEL allele frequency of other population, showed no major difference. The proportion of human genetic variation due to differences between populations is modest, and individuals from different populations can be genetically more similar than individuals from the same population. However, with sufficient genetic data allows correct classification of individuals into populations (Witherspoon et al., 2007).

4.3.2 Heterozygosity and HWE

Observed heterozygosity and expected across the 30 loci are presented in Table 4-2. (Seong et al., 2014) reported that low heterozygosity (below than 0.100) at HLD 118 was observed similar in two Asian populations (Larue et al. 2012, Fondevila et al. 2012). Across the Malaysian population, only in M-Chinese the observed heterozygosity for HLD 118 is 0.100. The mean observed heterozygosity for the Malay, M-Chinese, M-Indian, Iban and Bidayuh are 0.413, 0.376, 0.438, 0.394 and 0.379 respectively. The number of observed departures from HWE is no more than would be expected by chance. Overall the Malaysian populations have average of heterozygosity which indicate moderate diversity within the population. This also shows that the markers are

highly polymorphic by nature. The mean observed heterozygosity for the South African populations (Hefke et al., 2015) and Europeans (Martin et al., 2013) are about the same range as in the Malaysian populations.

None of the tested 30 loci at Malaysian population showed significant departure for HWE (after Bonferroni test), therefore they are statistically independent. HWE for locus HLD97 in the Mixed Ancestry, Indian Asian, Xhosa and Zulu population still deviated even after Bonferroni correction.

4.3.3 Linkage Disequilibrium

As for pairwise linkage disequilibrium (LD), there were no significant *p*-value after Bonferroni's correction. The absence of LD between all 30 INDEL loci, ensures the stability of the calculations concerning genotype frequency for forensic purposes. For Investigator DIPplex kit, no linkage had been detected from any published population data, However, there are several linked loci presented in commercial STR multiplexes which showed in Table 4-12 below. In this type of case, might potentially affects the estimation of close relatives (e.g. full brothers or second degree relatives) (Tamura et al., 2015) and ancestry analysis as product rule cannot be applied to get useful statistical calculations.

Table 4-12: Examples of loci located on the same chromosome and the distance apart [(Phillips et al. 2012) which presented in commercial STR multiplexes [adapted from Bright, Curran & Buckleton 2013, Bright et al., 2014)].

Locus pairs	Chromosomal location	Distance (cM)	Reference
vWA	12	11.9	(O'Connor et al., 2011)
D12S391			
D5S818	5	27.8	(Buckleton, Triggs 2006)
CSF1PO			
D21S11	21	44.7	(Buckleton, Triggs 2006)
Penta D			
TPOX	2	88.8	(Tamura et al., 2015)
D2S441			
D6S1043	6	4.4	(Tamura et al., 2015)
SE33			

4.3.4 F_{ST} and F_{IS}

The variance of pairwise FST value from one polymorphism to another among populations is related to determine whether that population experience i.e. genetic drift, migration and mutation. F_{ST} analysis was estimated using Arlequin (Excoffier, Lischer 2010, Excoffier et al., 2007) with genetic distance being estimated using Slatkins method (Slatkin 1995) with 10000 permutations applied for estimation significance. Meanwhile, FIS is called the inbreeding coefficient, and can be interpreted as a measure of inbreeding within the population.

Structural analysis indicated that although all ethnic groups are genetically closed to each other, yet they are substantially distinct**.** But the 30 INDEL marker used in this research did not show much difference. The low pairwise Fst value between Iban and Bidayuh showed their close relatedness which is simply explained by their common ancestors have migrated from Kalimantan (Chang et. al., 2009). In contrast, from the results obtained by Y-chromosome study (Chang et.al., 2009) revealed that the indigenous populations in Sarawak (Iban, Bidayuh and Melanau) are distinctly from each other, also to the three major ethnic groups in Malaysia (Malays, Chinese and M-M-Indian). Chinese and M-M-Indian showed slight higher pairwise FST values in any of the combinations with other populations which indicated genetic divergence, due to their distant original geographical locations (Tan 2001).

Beside to determine differentiation among populations, there are several advantages of the exact test method. Firstly, the test is unbiased and accurate even on small sample or with low frequency, secondly to provide test results for each locus and lastly there independent at ploidy level (Raymond, Rousset 1995). Between the Malaysian subpopulations, the M-Chinese, M-Indian and M-Indian, Iban pairs showed most differentiated populations and no difference at Chinese-Bidayuh pair except at locus HLD58 were seen.

In this study also, the exact test of population differentiations was carried out on 13 other worldwide populations; Han China, Korean, Uruguayan, Central Spain, Basque Country, Portugal, Zulu, Xhosa, Afrikaners, Indian South African, Mixed, Finland and Somalia, against the Malays, Chinese, M-Indian, Iban and Bidayuh respectively (see Appendix). As expected, after Bonferroni correction $p<0.0028$, minor or no significant differences were found between the Chinese and Han Chinese and Korean pairs respectively. On the other hand, between Malay, Iban and Bidayuh with the Africans ethnic groups (Zulu,

Xhosa, Afrikners and Somalians) pairs showed a relatively high distance at more than 18 INDEL loci. The M-Indian have a modest differentiation with the European group (Central Spain, Basque Country, Portugal and Finland) between 10 to 15 loci. Even though from the results showed that the DIPplex® Kit has a potential to estimate the geographically history, it has been reported that a larger marker panel more reliable to distinguish populations within a continent (Listman et al., 2007).

Based on the FST results, the values did not show large significant among the sub populations, indication of close genetic similarities. Thus, the FIS are expected to be close to 0. When FIS > 0, inferring that mating among the same ancestry randomly is predominant, or the population is partitioned into subpopulations and mating is more or less restricted within each subpopulation. In contrast, in the population with FIS < 0, avoidance of inbreeding or mating between subpopulations is carried out predominantly {(Nomura et al., 2001). The average F_{IS} value for Malaysian population has been calculated as 0.032 which is near to 0, indicated high level of random mating within the population and that the population is in random mating with a high level of heterozygosity.

In fact, as a multi-cultural country, it is a norm to practice mixed marriages among the sub populations. In fact, intermarriage has long been practiced in history, particularly to strengthen ties between the countries (Saat 2009). Therefore, the plural community in Malaysia was due to mixed marriages since then and now, which affected the population studies which be carried out in this chapter.

4.3.5 Population assignment of individual genotypes and AIMs

The HUGO Pan-Asian SNP Consortium reported that the East Asians and Southeast Asians shared a common origin, whilst the Central-South Asians shared ancestry with the Europeans (HUGO Pan-Asian SNP Consortium 2009). Furthermore, findings revealed that after analysing approximately 50,000 autosomal SNPs there are linked between the Malays, Chinese, Indonesian and the Indigenous people within the phylogenetic tree (Hatin et al., 2011).

The pentagonal (refer Figure 4-3) shows that the PCA of Malaysian subpopulations were dispersed. Thus indicated that the population can still differentiate using the INDELs. However, the M-Indian are the most divergent compared to the others. PCA for Malays, M-Chinese, Iban and Bidayuh were overlapping each other demonstrating there were

genetic similarities among them. Then, when the combination of the Malays, M-Chinese and M-Indian was analysed using the Snipper app, a triangular plot was obtained.

Based on the 52 SNP markers (Sharizah A. 2013) results showed that the Malays are more closely related to the Chinese, compared to the Malays with the M-Indian or the Chinese with the M-Indian. In contrast INDEL markers also gave a similar result but, the PCA for the Malays were distributed evenly towards the M-Chinese and M-Indian, indicated that the Malays are close to both. There was significant difference between the M-Chinese and M-Indian. Different outcomes obtained from the YSTR studied (Chang et al., 2007) on the Malay, M-Chinese and M-Indian. The M-Indian which is a divergent population yet somehow closer to Malay than M-Chinese.

The genetic differentiation among populations is very important to understanding human diversity and its historical origins. In general, the populations which were carried out in this study are closely related to be properly differentiated with small sets of 30 autosomal INDEL markers. The high heterogeneity of the Malaysian population from within and among populations, which justifies the need for additional studies involving more markers in the population to allow an accurate assessment of the genetic ancestry for each subpopulation. The 48-plex AIM INDEL (Santos et al., 2010) set can be used in the future as there are designed for broad-scale continental comparisons. Hence, the Qiagen Investigator® DIPplex kit are not well suited to study of geographically close populations as it was developed for forensic purposes.

4.3.6 Forensic indices

All INDEL loci are polymorphic in the population and the combined match probability (CMP) are 1 in 5.25×10^{-12} for Malay, 1 in 2.70×10^{-11} for M-Chinese, 1 in 1.79×10^{-12} for M-Indian, 1 in 4.87×10^{-11} for Iban, and 1 in 7.6×10^{-11} for Bidayuh. The combined power of discrimination is greater than 99.9999% and the combined power of exclusion are all 99.99%. A high CPD and low CMP indicate a low likelihood of random match confirming that the Qiagen Investigator® DIPplex kit is effective for identification of individuals in Malaysian population (Akhteruzzaman et al., 2013). The kit is shown to be highly polymorphic for all populations. In comparison, the combined indices (combined power of discrimination, combined match probability and combined power of exclusion) for other populations using the Qiagen Investigator DIPplex® Kit is summarized in Table 4-13 below. Generally, the results showed that the INDEL multiplex is suitable for forensic tool.

Table 4-13: Forensic parameters for the studied Malaysian populations and other published data.

Population group	CPD	CMP	CPE	Reference
Malay	0.999999	7.20×10^{-12}	0.9999	In this study
M-Chinese	0.999999	2.70×10^{-11}	0.9999	In this study
M-Indian	0.999999	1.79×10^{-11}	0.9999	In this study
Iban	0.999999	4.87×10^{-11}	0.9999	In this study
Bidayuh	0.999999	7.6×10^{-11}	0.9999	In this study
African-American	N/A	1.43×10^{-11}	0.9999	(Larue et al., 2012)
Asian	N/A	7.62×10^{-12}	0.9999	(Larue et al., 2012)
Caucasian	N/A	3.65×10^{-13}	0.9999	(Larue et al., 2012)
Southwestern Hispanic	N/A	2.12×10^{-12}	0.9999	(Larue et al., 2012)
Finish	0.9999999999996	3.54×10^{-13}	0.9961	(Neuvonen et al., 2012)
Somali	0.999999999995	5.03×10^{-12}	0.9862	(Neuvonen et al., 2012)
Hungarian	0.99999999999825	1.75×10^{-12}	0.9967	(Kis et al., 2012)
Central Spain	0.999999999999	N/A	0.9985	(Martin et al., 2013)
Basque	0.999999999999	N/A	0.9970	(Martin et al., 2013)
Bangladeshi	0.999 9998	2.87×10^{-12}	0.9947	(Akhteruzzaman et al., 2013)
South Portugal	0.999999999999	5.58×10^{-13}	0.9982	(da Silva et al., 2013)
Poland	N/A	7.98×10^{-14}	0.9900	(Pepinski et al., 2013)
Taiwan	N/A	1.22×10^{-11}	0.9884	(Pepinski et al., 2013)
Czech	0.999999999999853	6.8×10^{-12}	NA	(Zidkova et al., 2013)
Southern Brazil	0.999999999999646	3.54×10^{-13}	0.9973	(Torres et al., 2014)
Southern Koreans	0.99999999995	2.84×10^{-11}	0.9880	(Seong et al., 2014)
Chinese Han	N/A	1.80×10^{-11}	0.9880	(Wang et al., 2014)
Chinese She	N/A	3.17×10^{-11}	0.9800	(Wang et al., 2014)
Afrikaner	0.9999999999996	4.25×10^{-13}	0.9983	(Hefke et al., 2015)
Asian M-M-Indian	0.999999999998	2.23×10^{-12}	0.9944	(Hefke et al.,2015)
Mixed Ancestry	0.9999999999992	8.50×10^{-13}	0.9938	(Hefke et al., 2015)
Xhosa	0.99999999996	3.55×10^{-11}	0.9838	(Hefke et al., 2015)
Zulu	0.99999999994	6.42×10^{-11}	0.9783	(Hefke et al., 2015)
Iranian	N/A	3.33×10^{-13}	0.9970	(Poulsen L 2015)
Mexican	0.99999999	N/A	0.9863	(Martínez-Cortés et al., 2015)
Japanese	0.9999999998	2.67×10^{-11}	N/A	(Nunotani et al., 2015)
Greek	N/A	2.5×10^{11}	N/A	(Tomas et al., 2015)
Manila	0.99999999999997	N/A	0.9950	(Magalhães et al., 2015)
Iraq	N/A	3.00×10^{-12}	0.9970	(Tomas et al., 2016)
Turkey	N/A	3.00×10^{-12}	0.9970	(Tomas et al., 2016)

4.4 Conclusion

The bloodstained specimen samples used in this population study are the reference samples as they were used as database in the Department of Chemistry Malaysia (KIMIA).

Chapter 4 analyses INDEL profiles from large set of samples from five Malaysian populations (the Malay, M-Chinese, M-Indian, Iban and Bidayuh) to generate population genetic data for statistical evaluation of samples. The work estimates allele frequencies are similar across populations ranging from a inimum of 0.050 and maximum 0.950. The statistical analysis also determines all loci are in Hardy-Weinberg Equilibrium, therefore they are statistically independent. There are no significant Linkage Disequilibrium exists between markers which ensures the stability of the genotype frequency for forensic purposes, but there is significant differentiation between all populations (M-Indian are the most distinct from others), and some evidence of inbreeding within populations. The chapter also assessed a variety of forensic-related parameters for the populations, including the random match probability, polymorphic information content, power of discrimination and power or exclusion. The results show that the DIPplex markers meet forensic purposes.

CHAPTER 5

DEVELOPMENT, OPTIMISATION AND VALIDATION OF A NEW MULTIPLEX PCR ASSAY

In this chapter, two members of the Forensic Genetics Group, UCLAN that includes MRes student, Balnd Mustafa Albarzinji, from April 2014 to September 2015 and myself, jointly undertook a project.

5.1 Overview

Biological samples, particularly evidence samples are often exposed to unfavourable environmental conditions prior to collection. This can lead to DNA degradation to varying extents, which results in DNA fragmentation, leading to incomplete profiles. Furthermore, some samples give negative results due to PCR inhibitors, which can originate from the sample itself or might be introduced during sample processing or DNA extraction (Schrader et al., 2012). The problem continues when attempting to profile low quality DNA templates, which to separate between samples that are producing no or partial profiles because of DNA degradation and those that produce no or incomplete profiles because of PCR inhibition.

For such samples, it would be useful to determine the probability of degradation according to amplicon size, and available amount of DNA. However, development of multiplex success rates can be obtained by employing methods that assess both the quality of DNA (e.g. inhibition) and extent of DNA degradation prior to DNA typing. This method could potentially save time and money, by avoiding samples to be re-analyzed. Particularly, it is suitable when only limited amount of DNA can be extracted from casework samples, for example it is currently being used in separate project where samples recovered from Iraqi mass graves are being typed. The key application of the method is the ability to triage samples prior to attempting to produce a full STR, mini-STR or INDEL profile.

Previously, studies developed a multiplex PCR assay to assess the level of DNA degradation and monitor PCR inhibitions; a quadraplex that amplified fragments ranging between 70 bp to 384 bp that were specific to rabbit, pig and human (Nazir 2012). These

had been combined (Nazir et al., 2013) with two internally amplified controls (IACs) of 90 bp and 410 bp (Zahra et al., 2011, Zahra, Goodwin 2016).

Chapter 5 also relates to the development of new mini-assays to assess the level of degradation and inhibition in DNA samples, in order to triage samples. Chapter 5 describes the optimisation of an assay containing four amplicons of varying length (50 bp, 70 bp, 112 bp and 154 bp), and two internal amplication controls (90 bp and 170 bp), including assessing the sensitivity and specificity of the assay, the impact of a variety of different PCR inhibitors at varying concentrations and the effect of varying levels of degradation.

5.1.1 Objectives

- To develop and optimise four new markers in a mini 4-plex PCR; the target length of amplicons approximately 50 bp, 70 bp, 112 bp and 154 bp.
- Prepare a serial dilution and artificial degradation samples using a human genome female control DNA and test the sensitivity of the multiplex.
- To develop two internal amplified controls (IACs) of 90 bp and 170 bp (IAC_{90} and IAC_{170}).
- To combine and optimise two developed PCR systems; Mini 4-plex and IACs.
- Prepare serial degradation samples of known PCR inhibitors to test the sensitivity of the multiplex to different inhibitors.

5.2 Methods and materials

5.2.1 Multiplex design

5.2.1.1 Design of mini 4-plex primers

Four pairs of primers of short amplicons were designed to amplify 50 bp, 70 bp, 112 bp and 154 bp. The nuclear recombination activation gene 1 (RAG-1) from chromosome 11 (locus 11p13) was used to generate the 112 bp and 154 bp amplicons. Since the RAG-2 co-amplified with RAG-1, short amplicons of 50 bp and 70 bp were generated from RAG-2. This RAG-1 and RAG-2 are involved in somatic (V(D)J) rearrangement of B- and T-cell lymphocytes, which is essential for the development of a normal immune system and its functions.

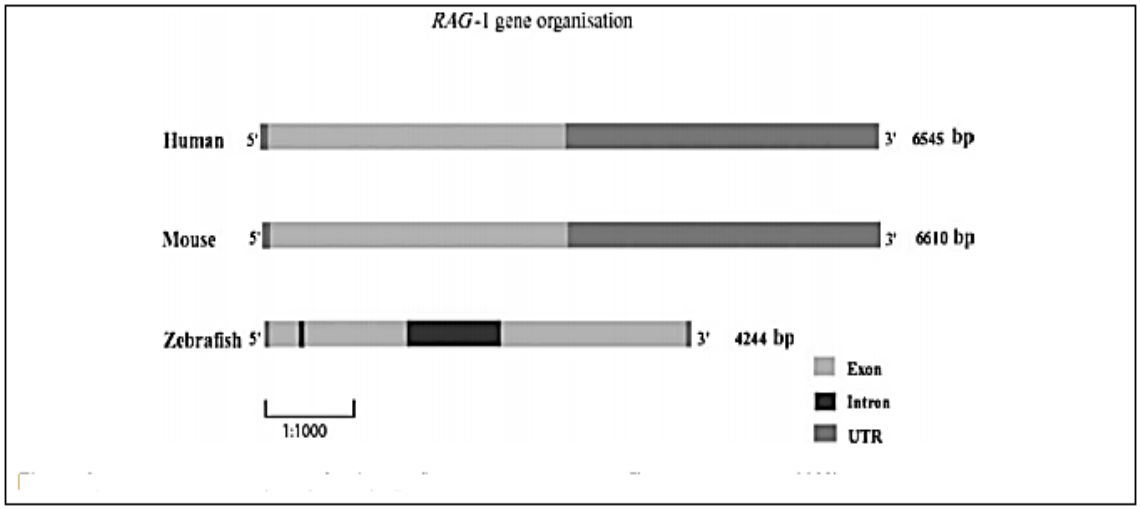


Figure 5-1**:** illustrates RAG-1 genes in human, mouse and zebrafish [adapted from (de Camargo, Nahum 2005)].

The sequence data for a nuclear recombination activation gene 1 (RAG-1) and gene 2 (RAG-2) from human and pig were downloaded from GenBank (http://www.ncbi.nlm.nih.gov/genbank/) and aligned using Bioedit Software V7.2 to identify conserved regions and the primers that would amplify 50 bp, 70 bp, 120 bp and 154 bp amplicons from the two species were identified.

Primer pairs with 5' fluorescein-labelled forward primers and unlabelled reverse primers were synthesized (Thermo Fisher Scientific) and purified using HPLC and desalting respectively. 100 µM stock solutions were prepared by adding the appropriate volume of distilled water (dH_2O) and stored at – 20 ºC, while an aliquot of 10 µM working solution was kept at 4 ºC.

5.2.1.2 Development of Internal Amplification Controls (IACs) fragments

The 90 bp and 410 bp fragments (IAC_{90} and IAC_{410}) were produced from plasmid pBR322 (Thermo Fisher Scientific) separately, using tailed primers (see Table 5-1) (Zahra et al. 2011). These fragments were generated from region 832 to 917 and 1682 to 2041 of the plasmid with composited primer technique as shown in Figure 5- 2.

Table 5-1: The sequence of the primers used to generate the IAC90 and IAC410 fragments.

Tailed primer	Sequence 5' – 3'
IAC_{90} forward	CTGTCAAATCTAAACACCCTGATGCGGCTTGCGGTATTCGGAATC TTG
IAC_{90} reverse	GTCAGCTTGCATAATATCGAGATAACGCGAGCGAGGGCGTGCAA GATT
IAC_{410} forward	TGTCAAATCTAAACACCCTGATGCGGATGCTGCTGGCTACCCTGT
IAC_{410} reverse	GTACAATGTTGACGTTCCTCGCTGCGTGAAGCGATTCACAGATCT CTG

Note: coloured letters show the sequence of the tailed primers.

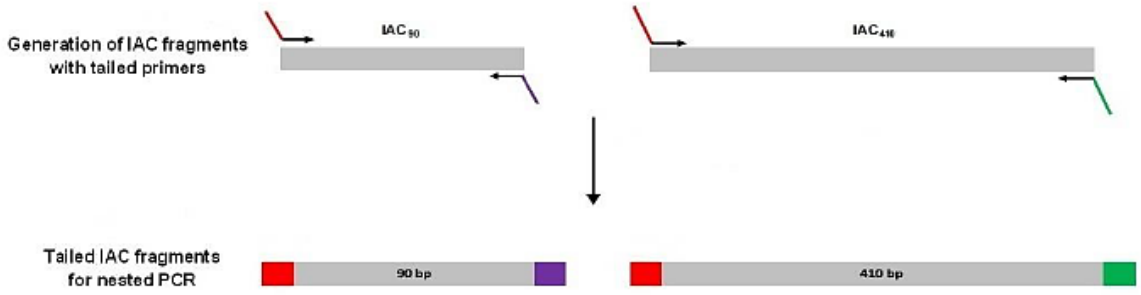


Figure 5-2: Schematic diagram shows the two steps PCR amplification of the plasmid pBR322 fragments using tailed primer (taken from (Zahra et al., 2011)) for the first generation of PCR.

Amplification reactions were prepared using 1 µl of 12 ng/µl plasmid, 6 µl of 2X ReddyMix PCR Master Mix (Thermo Fisher Scientific), 7 µl of nuclease-free water and 0.5 µl of 10 µM forward and reverse tailed primers to give a total reaction volume of 15 µl. Optimised thermal cycler conditions for IAC_{90} and IAC_{410} fragments are shown in Table 5-2. The PCR products were purified using QIAquick PCR Purification kit (Qiagen). The products were run on agarose gel electrophoresis. These fragments were diluted 10-fold until optimal concentration for balanced peaks obtained and kept as stock in -4 ºC.

Table 5-2: The PCR conditions for first generation of IAC_{90} and IAC_{410} from the plasmid pBR322 using Thermo Fisher Scientific 2X ReddyMix PCR Master Mix (1.5 mM $MgCl_2$).

PCR stage		Temperature (ºC)		Time
		IAC_{90}	IAC_{410}	
Initial incubation		95	95	2 min
Denaturing	35	95	95	25 s
Annealing		66	58	35 s
Extension		72	72	1 min
Final Incubation		72	72	5 min
Hold		4	4	∞

5.2.1.3 Nested PCR using Internal Amplification Controls (IACs) primers.

A slight modification was carried out on the primers pairs which used by (Zahra et al. 2011). This is because there was a gap obtained between the amplicon of 154 of the mini 4-plex and IAC_{410}. Therefore, a new reverse primer, IAC_{170} was designed to amplify a shorter fragment approximately 170 bp by using fluorescein labelled IAC_{410} forward primer. The IAC_{90} primer pair remained the same.

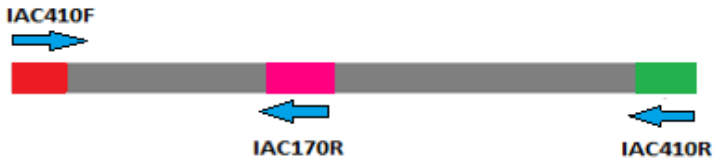


Figure 5-3: Schematic diagram shows the position of the new IAC_{170} reverse primer; to amplify the ROX labelled forward primer of IAC_{410} in the second PCR.

The amplification of IAC_{90} and IAC_{170} were carried out respectively using Scientific 2X ReddyMix PCR Master Mix (1.5 mM $MgCl_2$) together with labelled forward and unlabelled reverse primers (see Table 5-2. The thermal cycler conditions for amplification of IAC_{90} and IAC_{170} is as shown in Table 2-14, section 2.7.4.3. The reaction performed using Veriti® 96-Well Thermal Cycler (Thermo Fisher Scientific) and the PCR products were run on agarose gel electrophoresis.

Table 5-3: The sequence of the IAC_{90} and IAC_{170} ROX-labelled primers used in nested PCR.

Amplicon length (bp)	Forward and reverse primer (5' – 3')
90	CTGTCAAATCTAAACACCCTGATGCG
	GTCAGCTTGCATAATATCGAGATAACGC
170	CTGTCAAATCTAAACACCCTGATGCG
	GATGAACATGCCCGGTTACTG

5.2.3 Optimisation and sensitivity studies of Mini 4-plex PCR assay

Multiplex Mini 4-plex optimisation was carried out with primer concentration adjustment and experimental performance testing in an effort to generate maximal sensitivity, balanced peak heights and specific signals for four primer pairs in a single reaction. Primers with different concentrations mixed in a single reaction (see Table 5-4) and optimised using different temperatures (56 ºC to 62 ºC), PCR conditions and cycling parameters (see Table 5-5). In a total reaction volume of 10 µl PCR (5 µl Platinum® master mix (Thermo Fisher Scientific), 1 µl of IACs template, 0.6 µl of primer mix (forward primers 5´ fluorescein labelled), 2.4 µl of dH_2O and 1 µl of diluted Human Genomic Female DNA (0.5 ng) was used.

Table 5-4: The four primers sets for mini 4-plex PCR reaction tested for their optimum primer concentrations.

Primer		Annealing temperature (ºC)	Primer concentration (µM)				Amplicon length (bp)
			Set 1	Set 2	Set 3	Set 4	
50	F	60	0.060	0.062	0.064	0.060	50
	R	60	0.060	0.062	0.064	0.060	
70	F	60	0.065	0.068	0.067	0.067	70
	R	60	0.065	0.068	0.067	0.067	
112	F	60	0.060	0.050	0.054	0.054	112
	R	60	0.060	0.050	0.054	0.054	
154	F	60	0.065	0.068	0.064	0.067	154
	R	60	0.065	0.068	0.064	0.067	

Table 5-5: Thermal cycler conditions tested for the optimal multiplex PCR reaction amplification

PCR stage		Temperature (ºC)	Time (min)
Initial incubation		95	2
Denaturation	28 cycles	95	0.5
Annealing		56-62[(a)]	1.5
Extension		72	0.5
Final incubation		60	30
Hold		4	∞

[(a)]Every temperature from 56 to 62 ºC was tested separately with each set of the primer concentrations (Set 1 to Set 4).

5.2.4 New Mini 4-plex & IAC PCR system optimization

The new multiplex was developed using primers pairs of Mini 4-plex and two primer pairs of IACs. Thus, this multiplex amplifies (50, 70, 90, 112, 170 and 154) bp amplicons. The primer mix was prepared according to the optimised PCR condition (see Table 5-6).

Amplification of the Mini 4-plex and IACs was carried out using Platinum® Multiplex PCR Master Mix. The IACs purified fragments were diluted 10-fold until the concentration for balanced peaks obtained. A copy number (c.n.) of approximately 1120 for IAC_{90} and 886 c.n for IAC_{410} were mixed and used as IACs template in multiplex PCR. Different sets of primer mixture were prepared to optimise the peak heights of the markers (Mini 4-plex and the IACs). Human Genomic Female DNA (Promega) was diluted to 1 ng/µl and mixed with Platinum® master mix (5 µl), IACs (1 µl), primer mixture (0.6 µl) and nuclease-free water (2.4 µl) was added to adjust the reduced final volume of 10 µl.

Table 5-6: The concentration and volumes of the Mini 4-plex and the IACs primer mixture which prepared from 10 µM working solutions.

Primers	Forward and reverse primers	Primer volume	Primer concentration in a primer mix	Primer concentration in a 10 µl of PCR final volume
50	50 bp_ Fwd	10	1.06	0.063
	50 bp_Rev	10	1.06	0.063
70	70 bp_ Fwd	11	1.17	0.07
	70 bp_Rev	11	1.17	0.07
112	112 bp_ Fwd	10	1.06	0.063
	112 bp_Rev	10	1.06	0.063
154	154 bp_ Fwd	11	1.17	0.07
	154 bp_Rev	11	1.17	0.07
IAC_{90} & IAC_{170}	90 and 170 bp_Fwd	4	0.42	0.025
IAC_{90}	90 bp_ Rev	2	0.21	0.012
IAC_{170}	179 bp_Rev	3.8	0.4	0.024

Table 5-7: Thermal cycler conditions for Mini 4-plex & IAC PCR reaction amplification.

PCR stages		Temperature (°C)	Time (min)
Initial incubation		95	2
Denaturation	28 cycles	95	0.5
Annealing		60	1.5
Extension		72	1
Final incubation		60	30
Hold		4	∞

5.2.4.1 PCR inhibitors

The usage of the IACs in detecting the presence of PCR inhibitors were studied with varying concentrations of tannic acid, ethanol, phenol, 100X TE buffer (1 M Tris HCl, 0.1 M EDTA, pH 8.0), Dithiothreitol (DTT) (all from Sigma-Aldrich®) and CXR reference dye (Promega). The concentration ranges of each inhibitor tested is as shown in Table 5.8. A commercial female DNA control (193 ng/µl) (Promega) was used as the template DNA. The control was also amplified without any inhibitor as a blank control.

Table 5-8: PCR inhibitors and the concentrations used for PCR inhibitory study using multiplex system.

PCR inhibitor	Final concentration in a PCR reaction
Tannic Acid	(100, 50, 25, 12.5, 6.25, 3.12, 1.5, 0.7 and 0.3) ng
Ethanol	(9.6, 4.8, 3.5, 3, 2.5, 1.25, 0.15, 0.6, 0.3)%
Phenol	(99, 50, 25, 12.5 and 6.25) %
TE Buffer	(100, 50, 25, 12.5, 6.25 and 1.56) X
CXR dye	(30, 15, 7.5, 3.75 and 1.8) ng
DTT	(1, 0.5, 0.25, 0.125 and 0.06) M

5.3 Results

5.3.1 Development of Mini 4-plex multiplex PCR

5.3.1.1 Redesigning of primer 50 bp and 70 bp from RAG-2

At first, the primer all Mini 4-plex primers (50, 70, 112 and 154 bp) were designed from RAG-2. But, the mixture of the 4-plex primers caused problems; non-specific peaks which were amplified together (e.g. 374 bp and 384 bp). This due to the Mini 4-plex markers were closed to each other; 70 bp forward amplified with the reverse of 112 bp and forward of 112 bp amplified with the reverse of 50 bp which formed extra peaks. Since both RAG-1/2 co-expressed together, the 50 bp and 70 bp primers were redesigned on RAG-2, and the other two amplicons (112 bp and 154 bp) remained on RAG-1. These primers were used to form a new Mini 4-plex multiplex.

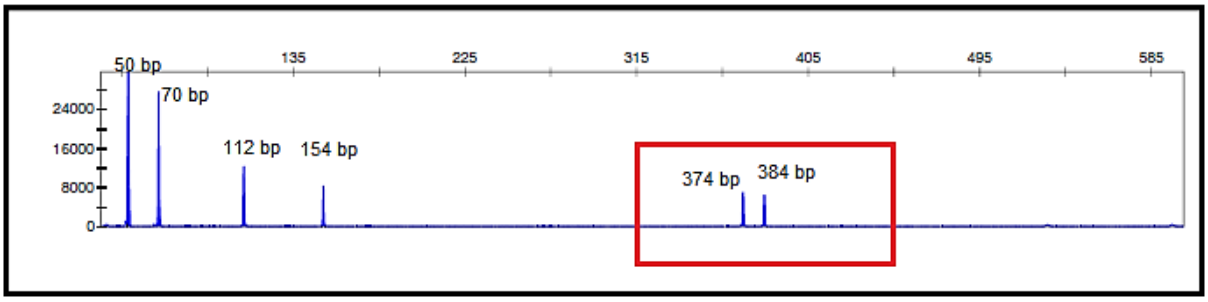


Figure 5-4: The electropherogram shows the non-specified peaks, which highlighted in the red box; the 112 bp primer amplified with the reverse/forward of 50 and 70 bp respectively.

5.3.2 Singleplex optimisation of the mini 4-plex markers

Amplification of the new markers were optimised using a ready-to-use Thermo Fisher Scientific 2X ReddyMix™ PCR Master Mix (1.5 mM $MgCl_2$) and Platinum® PCR SuperMix High Fidelity. PCR for each primer pair was set up individually using following PCR conditions: 94 °C for 3 min, followed by 94 °C for 1 min, 58 °C for 1 min, 72 °C for 1 min for 35 cycles and final extension at 72 °C for 10 min, and hold at 12 °C until samples were removed from the thermocycler.

Extracted DNA from buccal swabs and Promega Human Genomic Female DNA were used as a control in a PCR final volume of 15 µl using ReddyMix™ PCR Master Mix (Thermo Fisher Scientific) containing: 0.5 µl of each forward and reverse primer, 6 µl of ReddyMix™ PCR Master Mix, 1 µl of DNA template and 7 µl nuclease-free water.

The four primer pairs were found to be optimum at working concentration of 0.07 µM and at all temperatures (56, 58.60 and 62) ºC for (50, 70, 112 and 154) bp, respectively. The optimized primer pairs were assessed for any nonspecific amplification that would lead to extra peaks and could interfere with target peaks.

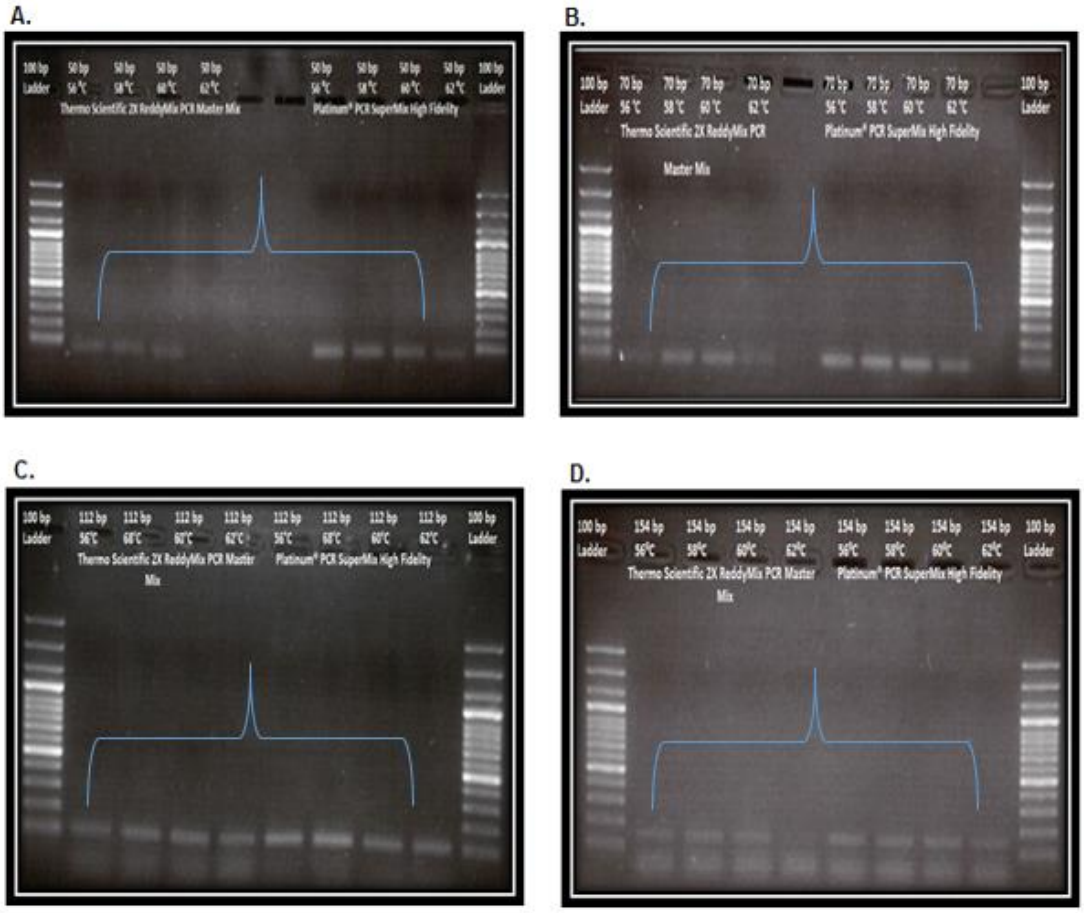


Figure 5-5: 2% gel electrophoresis (SafeView Nucleic Acid Stain) shows optimisation of the new markers individually (A) 50 bp, (B) 70 bp, (C) 112 bp and (D) 154 bp, using both 2X ReddyMix™ Master Mix and Platinum® PCR SuperMix High Fidelity with different annealing temperature (56, 58, 60 and 62 ºC).

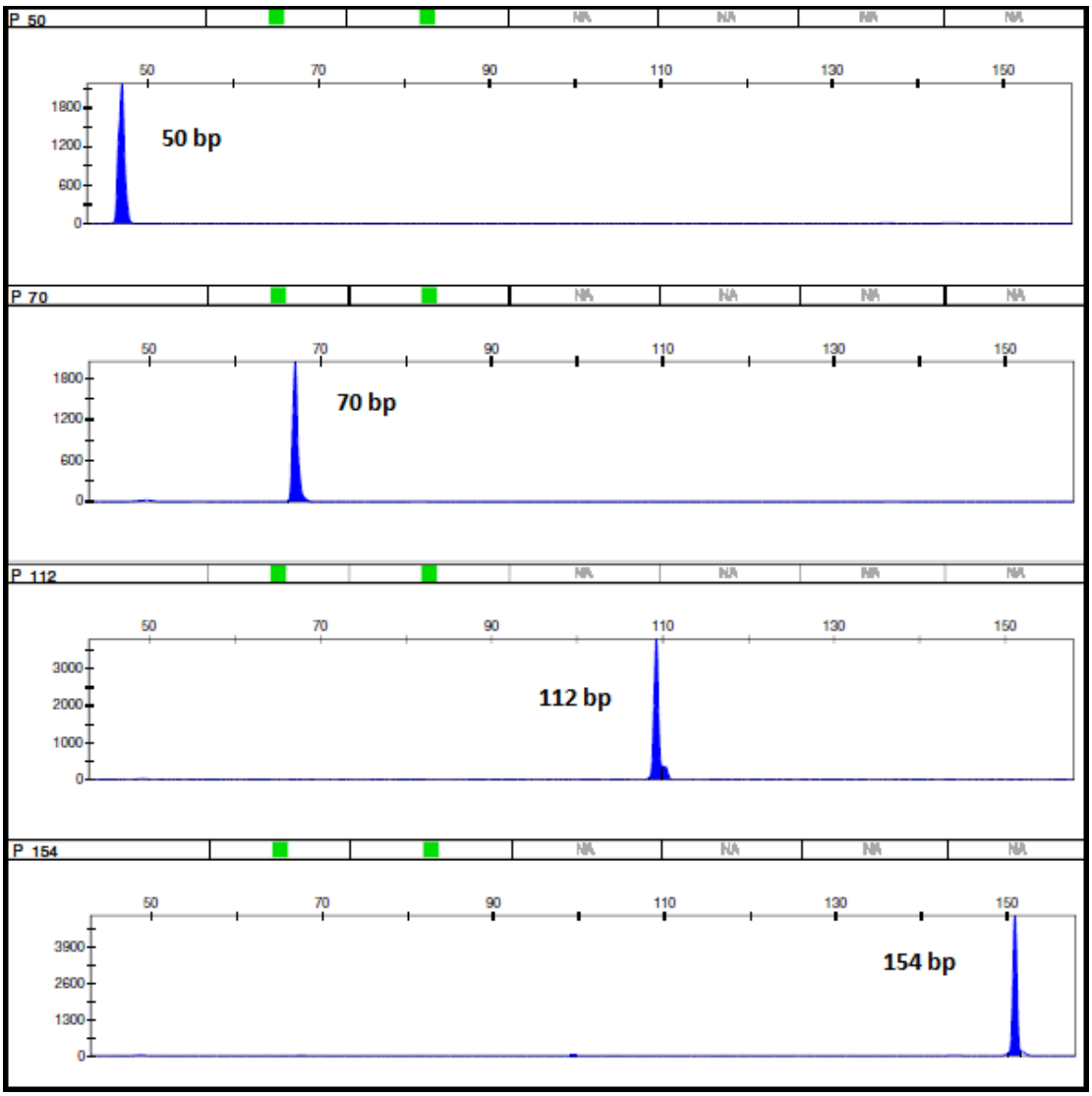


Figure 5-6: The electhropherogram of the singleplex of (50, 70, 112 and 154) bp. The amplification of the individual mini 4-plex markers was set up using the annealing temperature of 60 ºC for Platinum® PCR SuperMix High Fidelity (Thermo Fisher Scientific).

5.3.3 Multiplex amplification of human DNA samples using mini 4-plex markers

Amplification of the Mini 4-plex markers as a multiplex was performed using Platinum® Multiplex PCR Master Mix (Thermo Fisher Scientific). Extracted DNA from buccal swabs (4 ng/µl) and Human Genomic Female DNA (Promega) were used as a control template in a PCR reduced final volume of 10 µl.

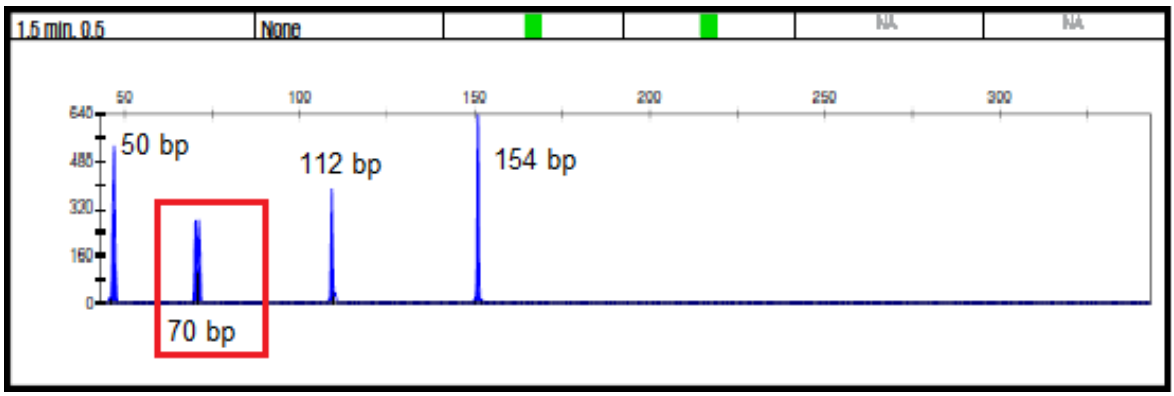


Figure 5-7: The imbalanced and split peak at 70 bp can be observed before the optimisation of the primer concentrations.

Primer mixture prepared as a separate reaction using different concentrations of the primers to balance the height of the peaks. Four different sets of primer concentrations were prepared to optimise and balance the mini 4-plex heights. Based on the results, the four primer pairs were found to be optimum at working concentration of (0.06, 0.065, 0.06 and 0.065) µM for (50, 70, 112 and 154) bp, respectively. The conditions of the primer set 1 was selected (Table 5-6). The amplification was carried out using the GeneAmp® PCR System 9700 Thermal Cycler. The PCR products were evaluated using the ABI 3500 Prism® Genetic Analyzer.

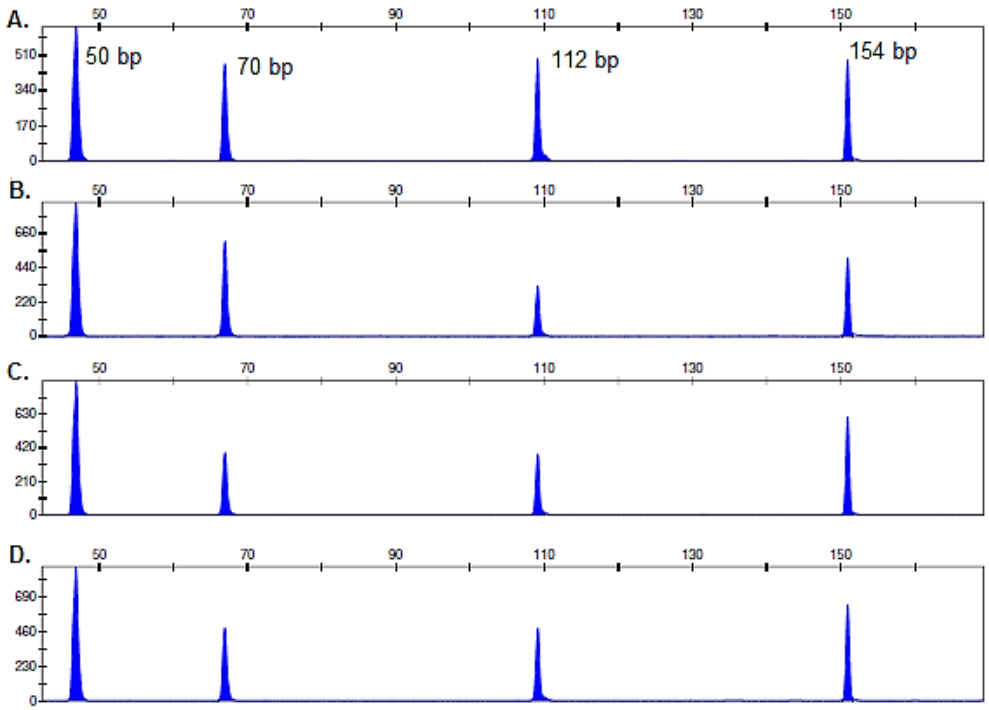


Figure 5-8: Electropherograms of (A) Set 1, (B) Set 2, (C) Set 3 and (D) Set 4. Each set consisted of different primer concentrations. Set 1 shows balanced peaks but imbalanced peaks can be observed in Set 2 to Set 4.

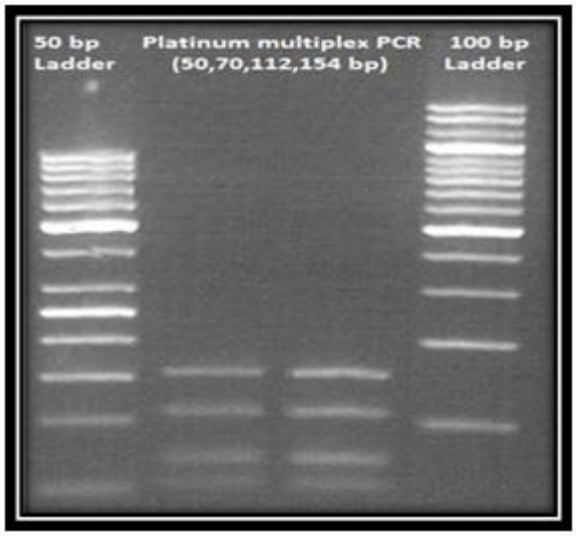


Figure 5-9: A 2% gel electrophoresis (GelRed Nucleic Acid Gel Stain) shows amplification of human DNA samples using the Mini 4-plex.

5.3.4 Amplicon specificity study

This assay was developed specific for human and pig (refer to 5.2.1.1.). So, one of the validation of this study was to determine the specificity of mini 4-plex to amplify human and pig samples in forensic cases. DNA from pig, roe, sheep, dog, rabbit and rat were amplified with Mini 4-plex (see Figure 5-10). A critical element of the PCR was that it should work with pig and also human DNA.

DNA from pig tissue was extracted using DNeasy® Blood and Tissue kit (Qiagen Ltd, UK) and the concentration was measured using Thermo Scientific NanoDrop Spectrophotometer 2000. The extracted samples for roe, sheep, dog, rabbit and rat were provided by Dr. Arati Iyengar (lecturer in School of Forensic and Applied Sciences). Amplifications were performed using Platinum® Multiplex PCR Master Mix in a final reaction of 10 µl containing: PCR master mix 5 µl, nuclease-free water 3.4 µl, non-human DNA template 1 µl and primer mix 0.6 µl of the optimised set shown in Table 5-4. Primer set 1 was used to simplify the samples. The amplification was performed using Veriti® 96-Well Thermal Cycler (Thermo Fisher Scientific) and the condition shown in Table 5-7. The PCR products were run on the 3500 Genetic Analyzer (Applied Biosystems).

Pig DNA sample tested produced a complete profile using the Mini 4-plex system. This shows that the multiplex worked also on pig samples. However, sheep displayed amplification products for (50, 70 and 154) bp. Roe displayed amplification products of

multiplex for 50 bp and 154 bp, and dog only at 50 bp. Furthermore, rabbit and rat did not display any amplification products for any of the markers within the Mini 4-plex.

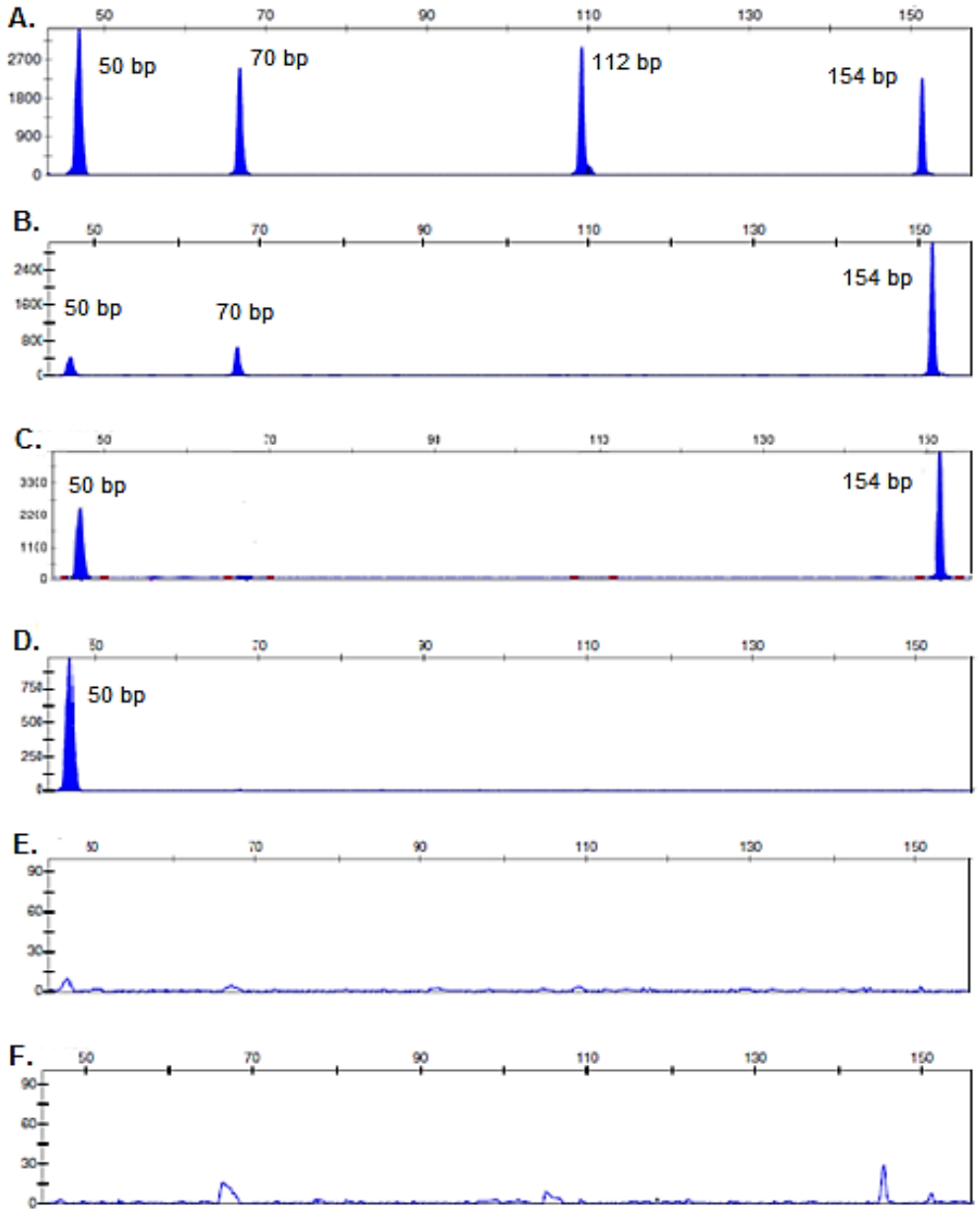


Figure 5-10: The amplification of (A) pig, (B) roe, (C) sheep, (D) dog, (E) rabbit and (F) rat using 1.0 ng of template. Different RFU scales y-axis are shown to demonstrate specificity. The profiles show that the Mini 4-plex successfully amplify pig sample but not for roe, sheep, dog, rabbit and rat. Therefore, this Mini 4-plex are specific to amplify human (refer to Figure 5-8) and pig DNA above.

5.3.5 Sensitivity study of Mini 4-plex markers

5.3.5.1 Mini 4-plex amplification of serial diluted DNA

A control DNA, Human Genomic DNA Female (Promega), was prepared by serial dilutions with different concentrations; (10, 5, 2.5, 1.25, 0.6, 0.3, 0.15, 0.07, 0.03, 0.01 and 0.009) ng and quantified using Real Time PCR. The GeneAmp® PCR System 9700 Thermal Cycler (Thermo Fisher Scientific) was used for amplification. The PCR products were sized and genotyped on the ABI 3500 Genetic Analyzer.

The results showed that full amplifications were observed from 10 ng down to 0.009 ng DNA. So, the minimum amount of DNA required for this multiplex to generate full profile is 0.009 ng (see Figure 5-11). A reduced peak height ratio of DNA 112 bp and 154 bp were seen at 0.3 ng onwards (see F, G, H, I, J and K in Figure 5-11). However, this is expected when amplifying less than optimal amount of DNA (0.5 ng/µl) due to stochastic effects. Whereas, the two robust and sensitive amplicons, 50 bp and 70 bp were remained above 50 RFU for all samples except at 0.01 ng and 0.009 ng respectively. Balanced peaks heights were developed at 0.6ng/µl to 2.5 ng/µl template DNA (see C, D and E in Figure 5-11). If excessive amount of DNA or amplification of the concentrated samples resulted in imbalanced peaks.

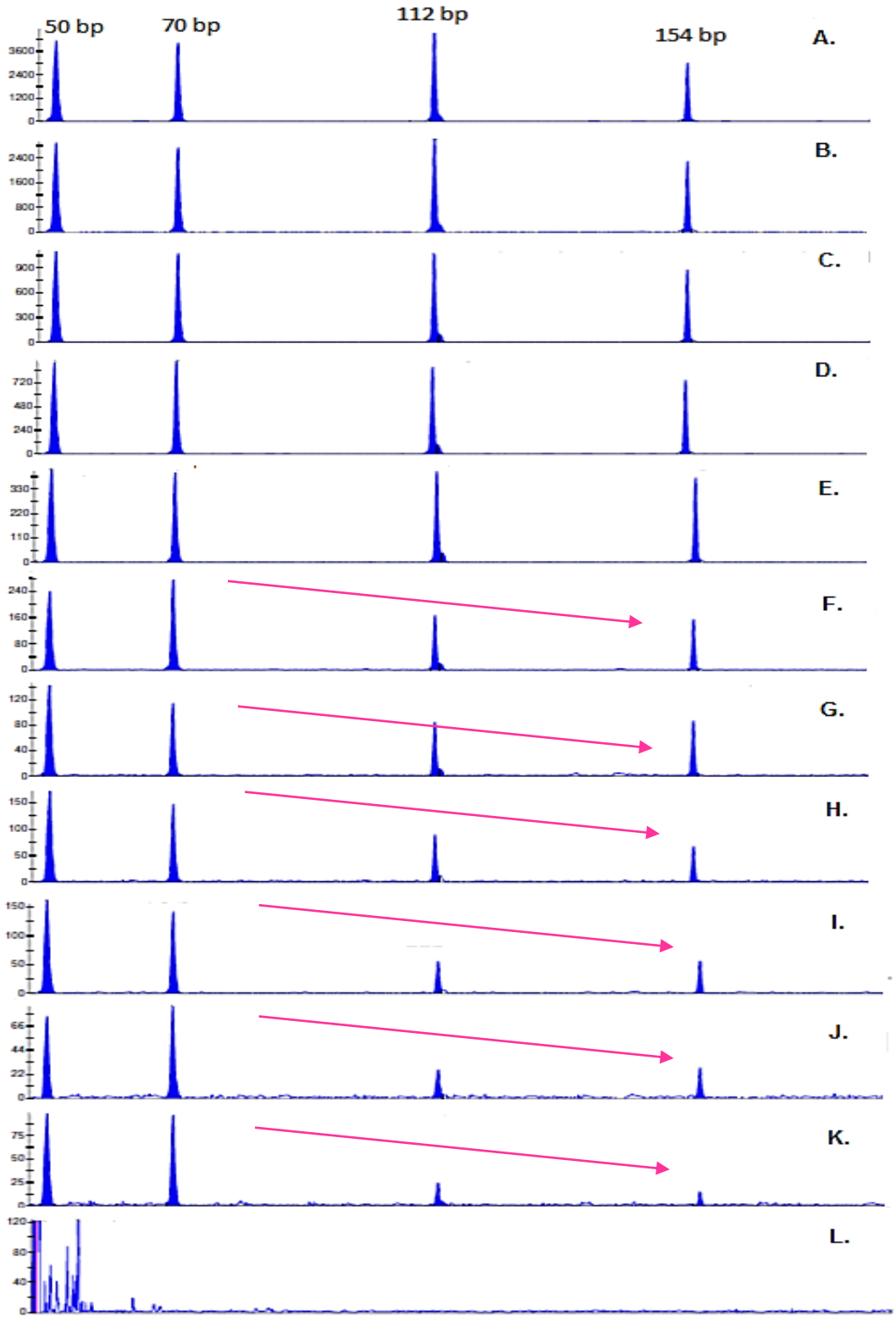


Figure 5-11: The effect of varying inputs of DNA template on peak height. The results above are representative of serial dilution of commercial female DNA (Promega) with concentrations (A) 10, (B) 5, (C) 2.5, (D) 1.25, (E) 0.6, (F) 0.3, (G) 0.15, (H) 0.07, (I) 0.03, (J) 0.01, (K) 0.009 ng/µl and (L) Negative Control. No DNA profile was developed from Negative Control.

5.3.5.2 Serial dilution samples with PowerPlex® 16 System and 4-plex PCR system.

In order to determine the efficacy of Mini 4-plex PCR assay on amplifying the extent on DNA degradation, amplification of a serial dilution of control DNA, Promega Human Genomic Female DNA was carried out. Concentrations of (10, 5, 2.5, 1.25, 0.6, 0.3, 0.15, 0.07, 0.03, 0.01 and 0.009) ng were amplified using the in-house 4-plex system (Nazir 2012). However, for the PowerPlex® 16 System (Promega), the concentration of 1.25, 0.6, 0.3, 0.15, 0.07, 0.03, 0.01 and 0.009 ng were used. The amplification was performed on GeneAmp® PCR System 9700 Thermal Cycler (Thermo Fisher Scientifc) using the PCR parameters as described in Chapter 2 and analysed on the ABI 3500 Genetic Analyzer. The minimum threshold used in both PCR multiplexes for peak detection was 50 RFU.

The results showed that in 4-plex multiplex (see Figure 5-12), full profiles were obtained from 10 ng/µl to 0.07 ng/µl DNA, thus the minimum amount of DNA required for this multiplex to generate full profile was 0.07 ng. Inconclusive results were generated with 0.03 and 0.01 ng/µl DNA as the RFU was under 50 and incomplete. No profile was developed with 0.009 ng/µl and negative control.

Full DNA profiles were developed using 0.3 ng to 1 ng in PowerPlex® 16 System (Promega) (see Figure 5-12). Below than that concentration, partial profiles were generated where allele drop out started to occur at larger loci, e.g. D18S51, Penta E, FGA and Penta D. Partial profiles can be obtained at 0.07 to 0.15 ng/µl. Primarily, the allele called in PowerPlex® 16 System (Promega) came from the smaller loci. However, weak and inconclusive or negative results were developed with DNA concentration below than 0.03 ng/µl.

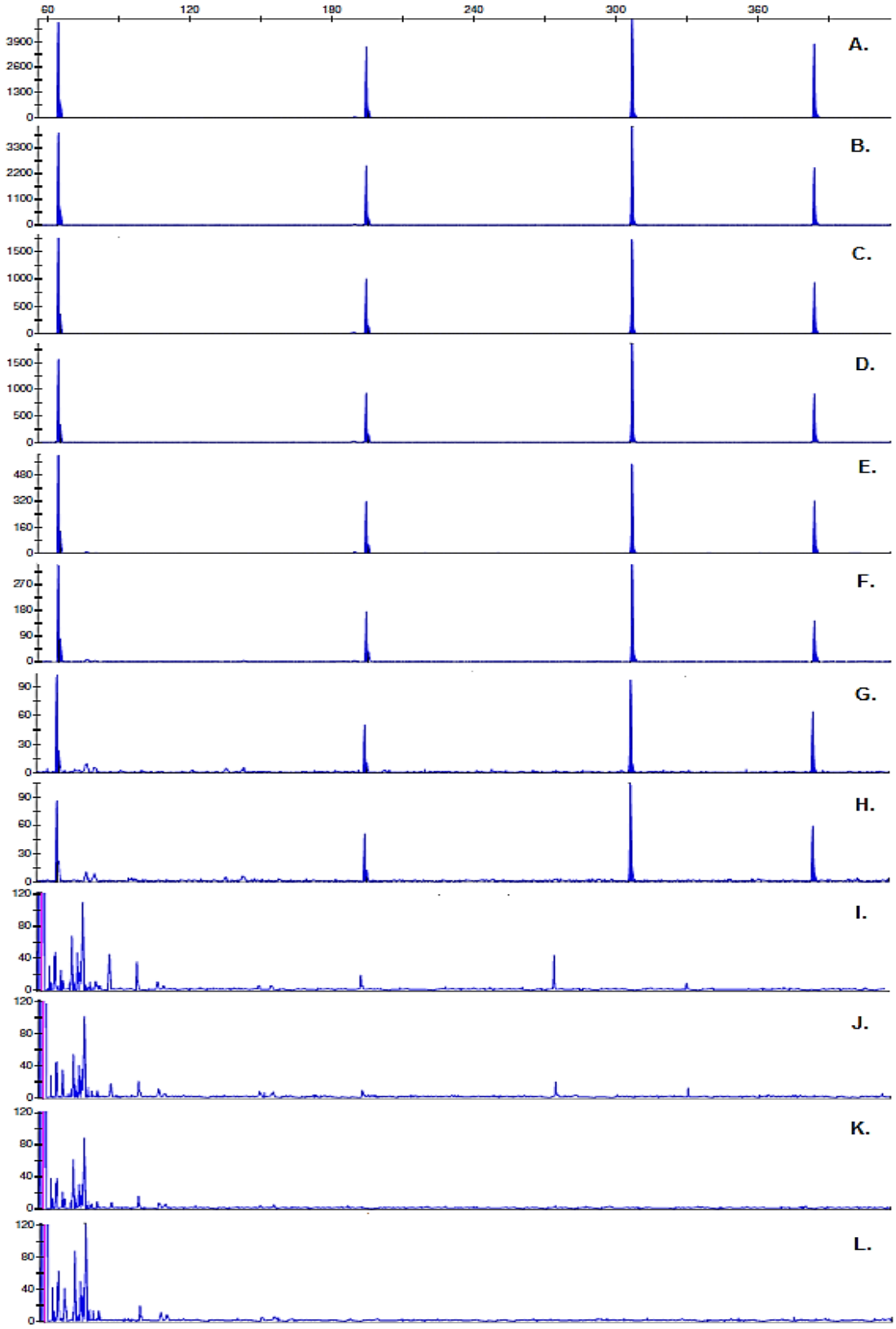


Figure 5-12: Electropherogram shows the 4-plex system amplification of serial dilution control DNA with concentrations of (A) 10, (B) 5, (C) 2.5, (D) 1.25, (E) 0.6, (F) 0.3, (G) 0.15, (H) 0.07, (I)

0.03, (J) 0.01, (K) 0.009 ng/µl and (L) Negative Control. No DNA profile was developed from Negative Control.

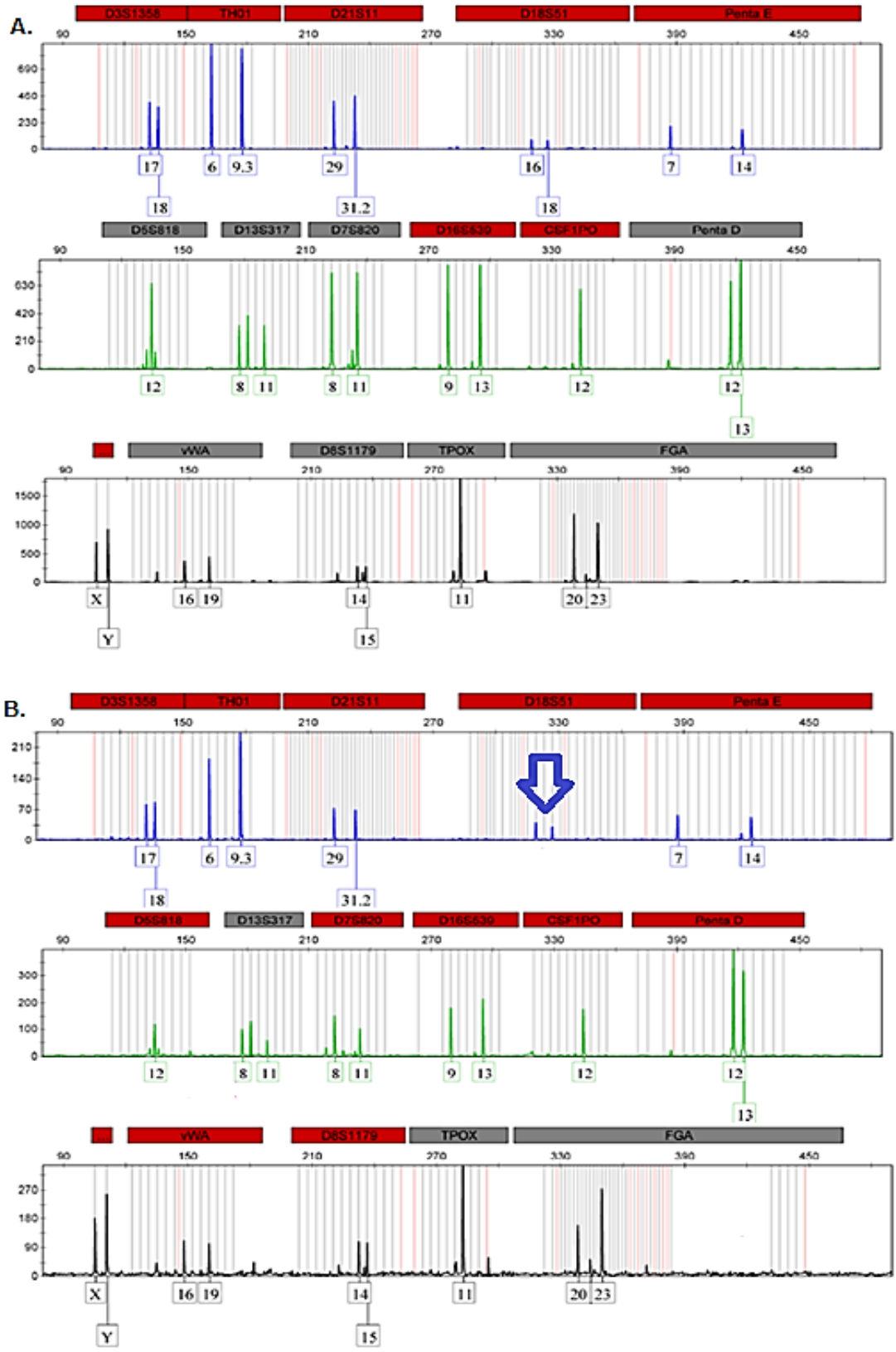


Figure 5-13: Electropherograms show PowerPlex®16 System amplification of (A) full DNA profile of 1.25 ng control (2800M) (refer to (D) for the 4-plex system in Figure 5-12). And (B) drop out allele D18S51 with amplification of diluted control (2800M) (0.15 ng/µl) (refer to (G) for the 4-plex system in Figure 5-12).

Table 5-9: The sensitivity study of the Mini 4-plex PCR assay compared to 4-plex system (Nazir et al. 2013, Nazir 2012) and PowerPlex® 16 System (Promega).

Concentration (ng)	Mini 4-plex PCR assay	4-plex system	PowerPlex® 16 System
1.25	++++	++++	++++
0.6	++++	++++	++++
0.3	++++	++++	++++
0.15	++++	++++	++--
0.07	++++	++++	++--
0.03	++++	----	----
0.01	++++	----	----
0.009	++++	----	----

5.3.5.3 Model of DNA degradation

Artificial degradation of the commercial Human Genomic Female DNA (Promega) was prepared to assess the ability of the Mini 4-plex to quantify DNA in degraded samples using Promega RQ1 RNase-Free DNase I and serial digestions were taken at intervals from 2 min to 180 min. The degree of degradation was assessed, the prepared samples were amplified and the products were analysed on the ABI 3500 Genetic Analyzer.

The results showed that the markers produced full Mini 4-plex profiles were obtained until 60 min of degradation with peak heights above 100 RFU; where the peak heights are proportional to the detected PCR products (see Figure 5-14). The 50 bp and 70 bp markers were amplified until 120 min of digestion approximately 100 RFU and 200 RFU respectively. As expected the 112 bp and 154 bp amplicons started to decrease the heights with increasing of digestion time. At 90 min and 120 min of degradation, it could

see both larger markers are amplified but below 50 RFU. No profile was developed at 180 min digestion.

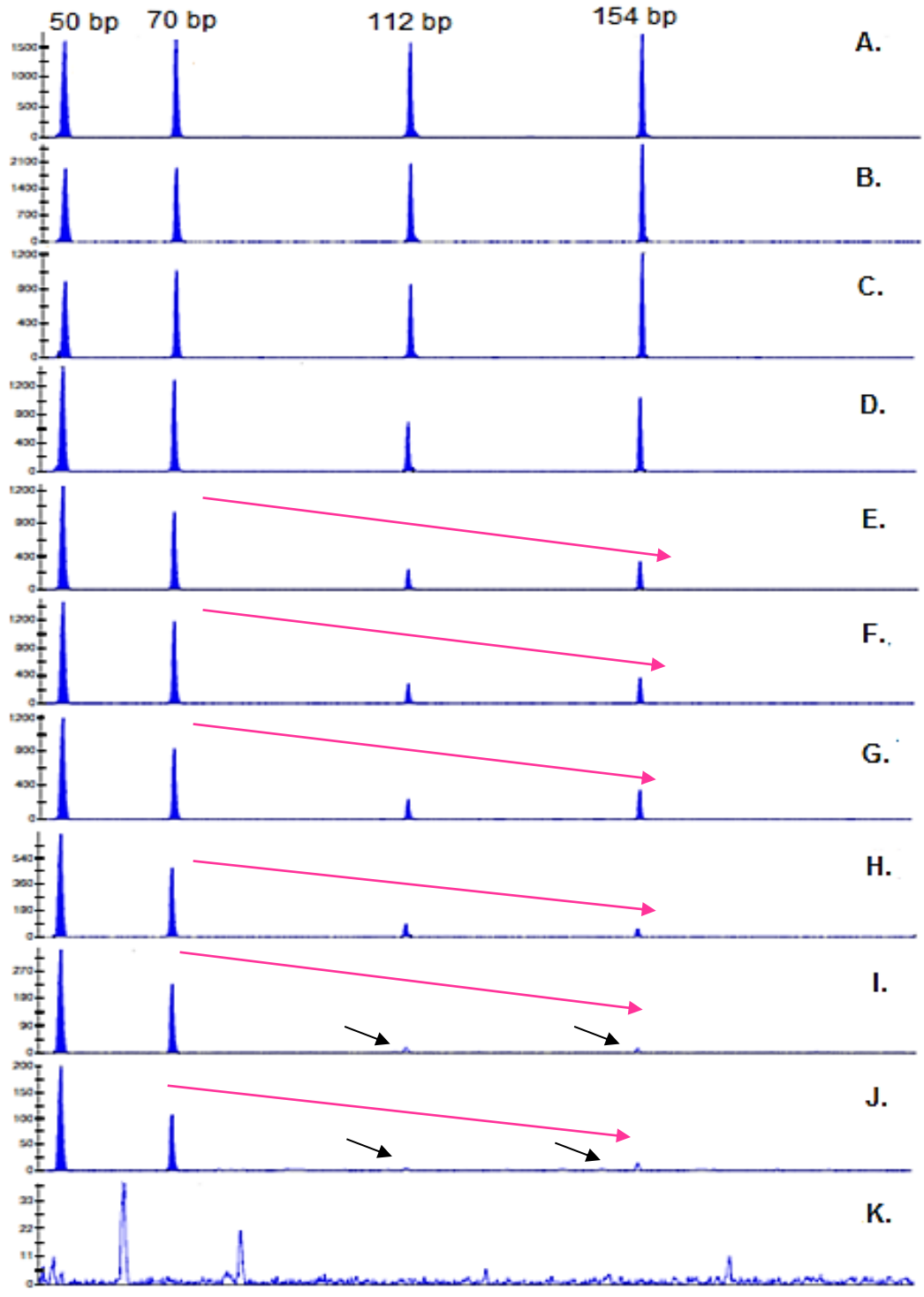


Figure 5-14: Electropherogram showing Mini 4-plex amplification of artificially degraded DNA using DNase I (100x diluted) at time of points (A) 0 min, (B) 2 min, (C) 5 min, (D) 10 min, (E) 20 min, (F) 30 min, (G) 45 min, (H) 60 min, (I) 90 min, (J) 120 min and (K) 180 min.

5.3.6 Combination of mini 4-plex and Internal Amplification Control (IACs)

In the early stages of this research, the mini 4-plex was combined with the IACs which were designed by previous student (Zahra, Hadi & Goodwin 2012); IAC_{70} and IAC_{410}. However, there was a gap between the amplicon 154 bp of the mini 4-plex and the IAC_{410} (not shown here). A modification in the reverse amplicon of IAC_{410} by reducing the target size to 170 bp. The IAC_{170} was used throughout this research.

The new multiplex was designed by adding four primer pairs that were designed to amplify the amplicons length of (50, 70, 112 and 154) bp for pig and human together with two primer pairs for Internal Amplification Controls (IACs) to amplify 90 bp and 170 bp amplicons. The G+C contents of each primer were between 40- 60 % and the primer length of all the primers was less than 25 nucleotides.

5.3.6.1 Amplicon specificity study

The optimised multiplex (Mini 4-plex & IACs) system was assessed for any non-specific amplification that would lead to extra peaks and could interfere with the target loci. A commercial female DNA control (Promega) was used as a template to evaluate the multiplex. The amplified products were analysed on ABI 3500 Genetic Analyzer according to the protocol described in Table 5-7. All six amplicons were amplified and produced specific peaks for the targeted loci without any additional peaks (Figure 5-15).

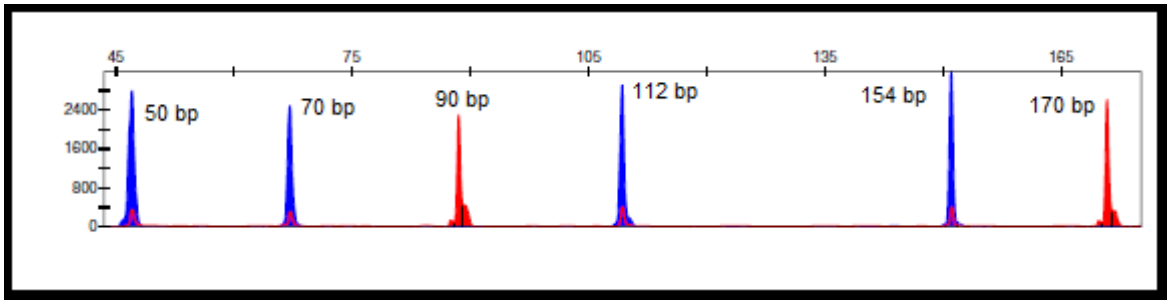


Figure 5-15: examples of electropherogram shows specific peak for each amplicon (50, 70, 90, 112, 154 and 170 bp) without any additional peak(s).

5.3.6.2 PCR inhibitors study using the new multiplex system

Amplifications were performed with 1 ng of Human Genomic Female DNA in 10µl Mini 4-plex & IACs multiplex reactions with increasing concentrations of various PCR inhibitors. 1 µl of different types of inhibitors (e.g. tannic acid, ethanol, TE buffer (10mM Tris-HCl pH 8, 1 mM EDTA), phenol, DTT (1,4-Dithiothreitol) (all from Sigma-Aldrich®) and CXR dye reference (Promega) were prepared to test the efficiency of the IACs in the multiplex system to detect the inhibitors. Each sample point was prepared in triplicate but only one sample was chosen to represent them in the ratio calculation since all the triplicate samples produced similar results. The control DNA was amplified together without any inhibitor as a blank control. The results are shown in Table 5-10 and Table 5-11.

An inhibitory effect was observed in all profile. As illustrated in Table 5-10 and Table 5-11, when the concentration of inhibitor increased, the quality of detectable decreased. The difference of peak height ratio was observed and calculated where the highest peak height (RFU) divided to the lowest one.

The results in Figure 5-16 showed that full profiles were generated at lower concentrations at 1.5 ng/µl and 3.12 ng/µl of tannic acid using the Mini 4-plex & IACs system. However, at 6.25 ng/µl tannic acid, the (50, 90, 112, 154) bp from the Mini 4-plex& IACs system callable peaks (see Figure 5-16), whereas at 12.5 ng/µl and 25 ng/µl there are dropout at all markers except for the robust 50 amplicon and IAC90 with peak heights above 50 RFU. No profile was developed at concentrated tannic acid; 50 and 100 ng/µl**.**

The electropherograms in Figure 5-17 showed that higher concentration (9.6% and 4.8%) of ethanol gave no profile. However, as the concentration reduced, full profiles were developed at 3.5%, 3%, 2.5% and 1.25%. Balanced peak heights obtained at 3% ethanol

The results obtained from co-amplification with TE buffer (see Figure 5-18) showed good quality profiles were developed from the concentration of 1.56X to 12.5 X. At 25X TE buffer, dropouts observed at all Mini 4-plex & IACs markers, except for IAC_{90}. The peak heights observed were above 50 RFU. For the maximum concentration of TE buffer, no profile was obtained (see Figure 5-18).

As some labs still extract DNA from casework stains using phenol-chloroform method, the effect of phenol on DNA was observed (see Figure 5-19). All amplicons had dropped out at 99 % and 50 % phenol. With 25 %, 12.5 % and 6.25 % phenol, Mini 4-plex & IACs system had alleles above 50 RFU at all markers.

Full profiles were detected in the presence of CXR dye up to 30 ng and 1 M of DTT per 1µl amplification reaction respectively with the Mini 4-pex & IACs system, which indicated no inhibitors detected in the profiles (see Figure 5-20 and Figure 5-21). Control amplification of Human Genomic Female DNA without inhibitors, did not show any inhibition of the Mini 4-plex & IACs system

Table 5-10: Table below shows the properties of DNA profiles developed with tannic acid, ethanol and TE buffer with different concentrations.

Inhibitor	Final conc	Mini 4-plex PH (RFU)				Peak imbalance	IACs PH (RFU)		IAC_{170}/ IAC_{90}
		50	70	112	154		IAC_{90}	IAC_{170}	
Tannic Acid	100 ng/µl	NP	NP	NP	NP	NP	NP	NP	NP
	50 ng/µl	NP	NP	NP	NP	NP	NP	NP	NP
	25 ng/µl	192	NP	NP	NP	NP	54	NP	NP
	12.5 ng/µl	215	NP	NP	NP	N/A	79	NP	N/A
	6.25 ng/µl	452	NP	344	113	PP	349	NP	N/A
	3.12 ng/µl	1199	749	1040	1162	0.62	546	434	0.85
	1.5 ng/µl	1654	1685	1767	1870	0.88	461	539	0.85
Ethanol	9.6%	NP	NP	NP	NP	NP	NP	NP	NP
	4.8%	NP	NP	NP	NP	NP	NP	NP	NP
	3.5%	745	403	785	753	0.51	440	427	0.90
	3.0 %	979	840	959	892	0.85	617	632	0.97
	2.5 %	688	660	819	752	0.80	587	673	0.87
	1.25 %	554	582	666	657	0.83	607	623	0.97
TE Buffer	100X	NP	NP	NP	NP	NP	NP	NP	NP
	50X	NP	NP	NP	NP	NP	NP	NP	NP
	25X	2126	2359	1554	1961	0.66	2309	3683	0.63
	12.5X	1900	2110	1473	1927	0.70	1875	3064	0.61
	6.25X	1956	2027	1532	1953	0.76	2188	3712	0.59
	1.56X	1704	1632	1397	1604	0.82	2147	3577	0.60

Note: N/A: Not Applicable, NP: No Profile, PP: Partial Profile.

Table 5-11: The properties of DNA profiles developed with tannic acid, ethanol and TE buffer with different concentrations and the control DNA

Inhibitor	Final conc	Mini 4-plex PH (RFU)				Peak imbalance	IACs PH (RFU)		IAC_{170}/ IAC_{90}
		50	70	112	154		IAC_{90}	IAC_{170}	
Phenol	99 %	NP	NP	NP	NP	NP	NP	NP	NP
	50%	NP	NP	NP	NP	NP	NP	NP	NP
	25%	2347	1964	2079	2488	0.79	2704	4566	0.59
	12.5%	2779	2677	2334	2546	0.83	2750	4541	0.61
	6.25%	2321	2208	2044	2560	0.86	2332	3950	0.59
CXR	30 ng/µl	2350	2401	1981	2550	0.77	2369	4112	0.58
	15 ng/µl	2406	2380	2062	2255	0.86	2323	3941	0.59
	7.5 ng/µl	1925	1879	1684	1762	0.87	2404	4003	0.60
	3.75 ng/µl	1820	1972	1567	1955	0.80	1885	3187	0.59
	1.8 ng/µl	2204	2323	1865	2456	0.76	2215	3858	0.57
DTT	1M	1912	922	1594	1584	0.48	2320	3931	0.59
	0.5M	1889	1625	1749	2123	0.77	2045	2530	0.58
	0.25M	2361	2412	2151	2787	0.77	2139	3726	0.57
	0.125M	2546	2461	2240	2261	0.89	2494	4137	0.60
	0.06M	2431	2235	2028	2305	0.95	2993	5037	0.59
Positive Control	1.0 ng/µl	2232	1998	2398	2482	0.80	2133	2633	0.81

Note: N/A: Not Applicable, NP: No Profile, PP: Partial Profile.

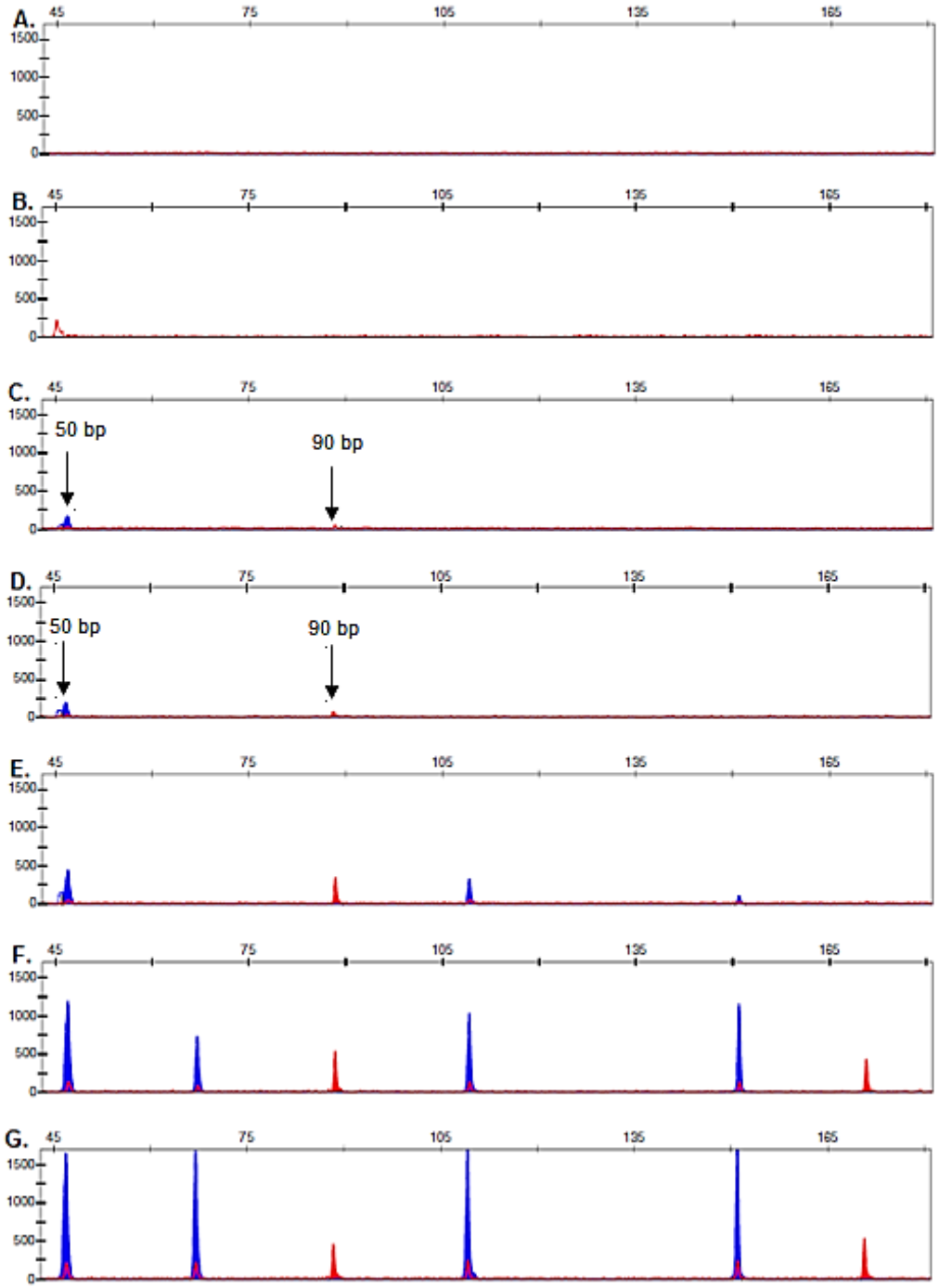


Figure 5-16: Electrophergrams generated from 1 ng of Human Genomic Female (Promega) control with final concentrations of A) 100 ng/µl, (B) 50 ng/µl, (C) 25 ng/µl, (D) 12.5 ng/µl, (E) 6.25 ng/µl (F) 3.12 ng/µl and (G) 1.5 ng/µl tannic acid in 10 µl PCR final volume. The blue peaks are the mini 4-plex which to detect degradation, while the red peaks are the IACs, to detect inhibition.

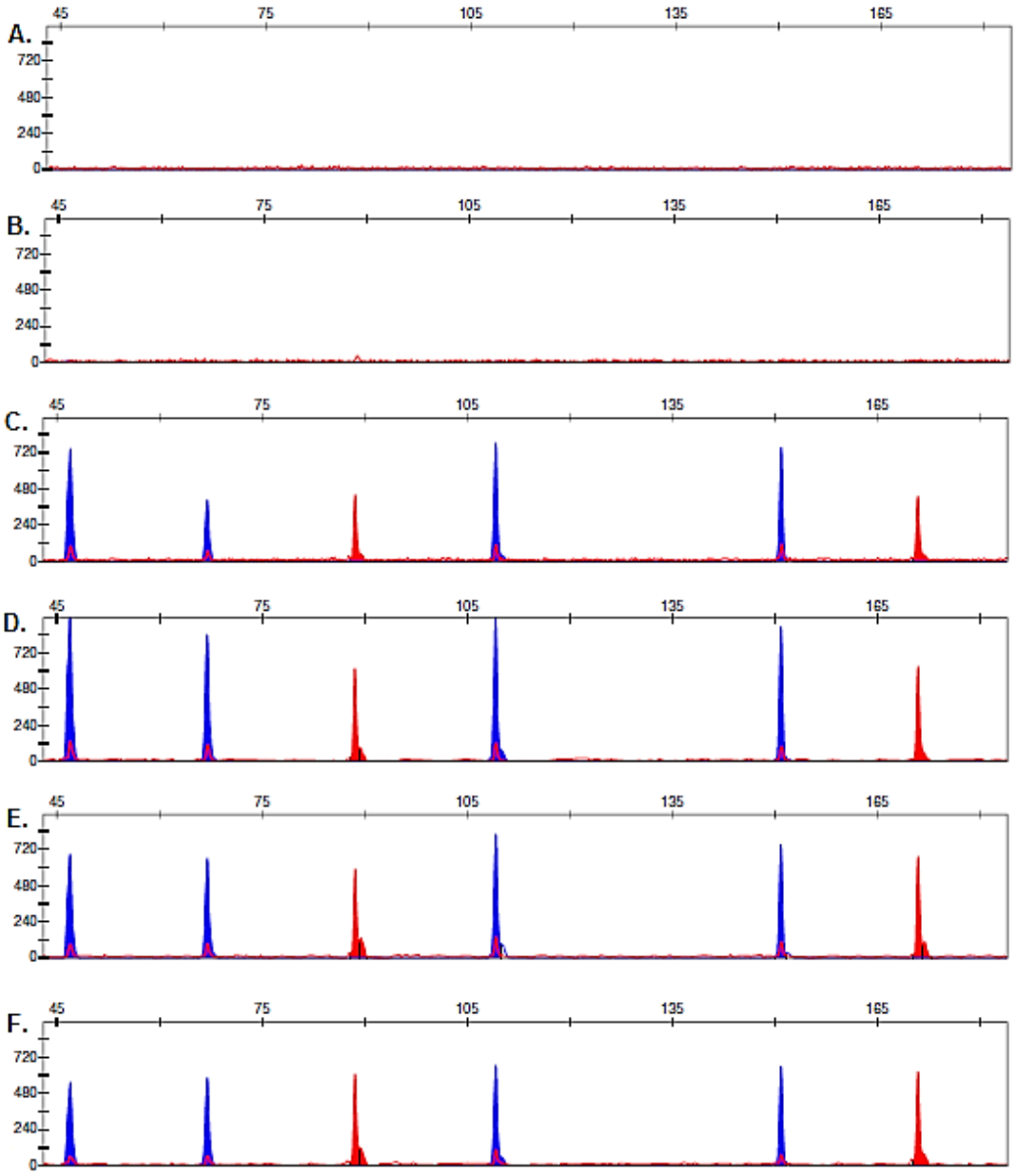


Figure 5-17**:** Electrophergrams generated from 1 ng of Human Genomic Female (Promega) control with final concentrations of (A) 9.6%, (B) 4.8%, (C) 3.5%, (D) 3%, (E) 2.5% and (F) 1.25 %Ethanol in 10 µl PCR final volume. The blue peaks are the mini 4-plex which to detect degradation, while the red peaks are the IACs, to detect inhibition..

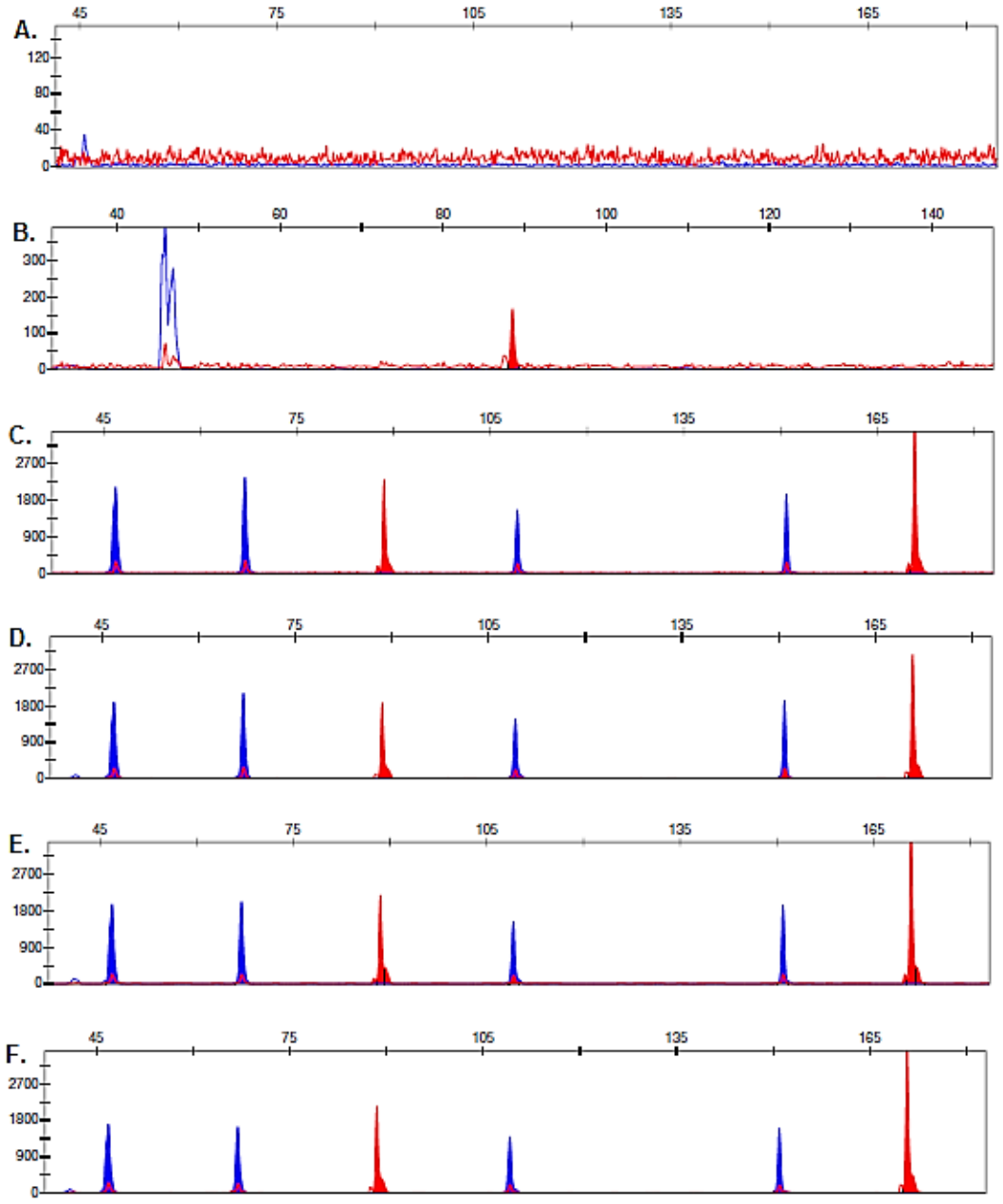


Figure 5-18**:** Electrophergrams generated from 1 ng of Human Genomic Female (Promega) control with final concentrations of (A) 100X (B) 50X, (C) 25X, (D) 12.5X, (E) 6.25X and (F) 1.56X TE buffer in 10 µl PCR final volume. The blue peaks are the mini 4-plex which to detect degradation, while the red peaks are the IACs, to detect inhibition.

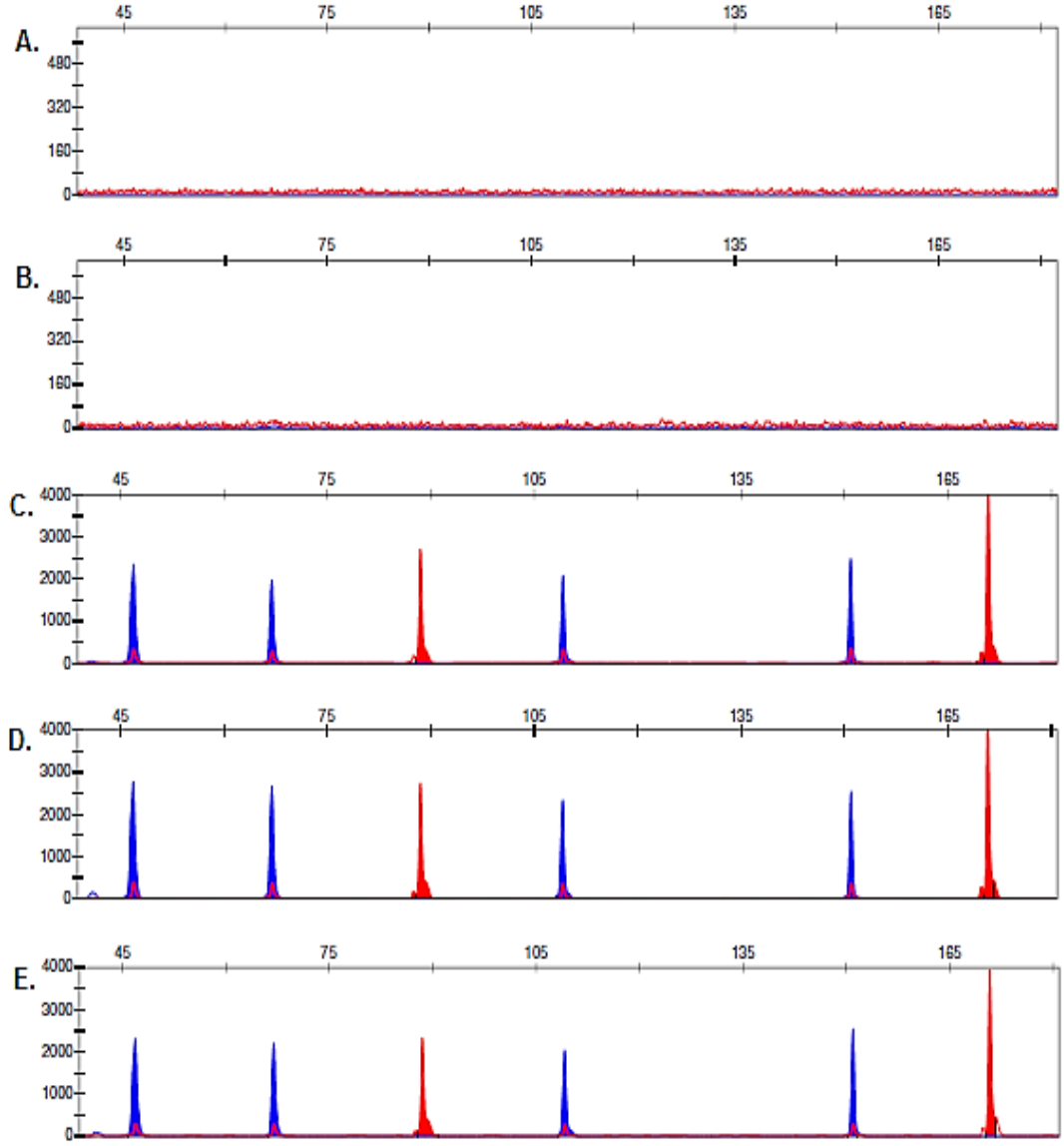


Figure 5-19**:** Electropherograms generated from 1 ng of Human Genomic Female (Promega) control with final concentrations of (A) 99% (B) 50%, (C) 25%, (D) 12.5% and (E) 6.25% Phenol in 10 µl PCR final volume. The blue peaks are the mini 4-plex which to detect degradation, while the red peaks are the IACs, to detect inhibition.

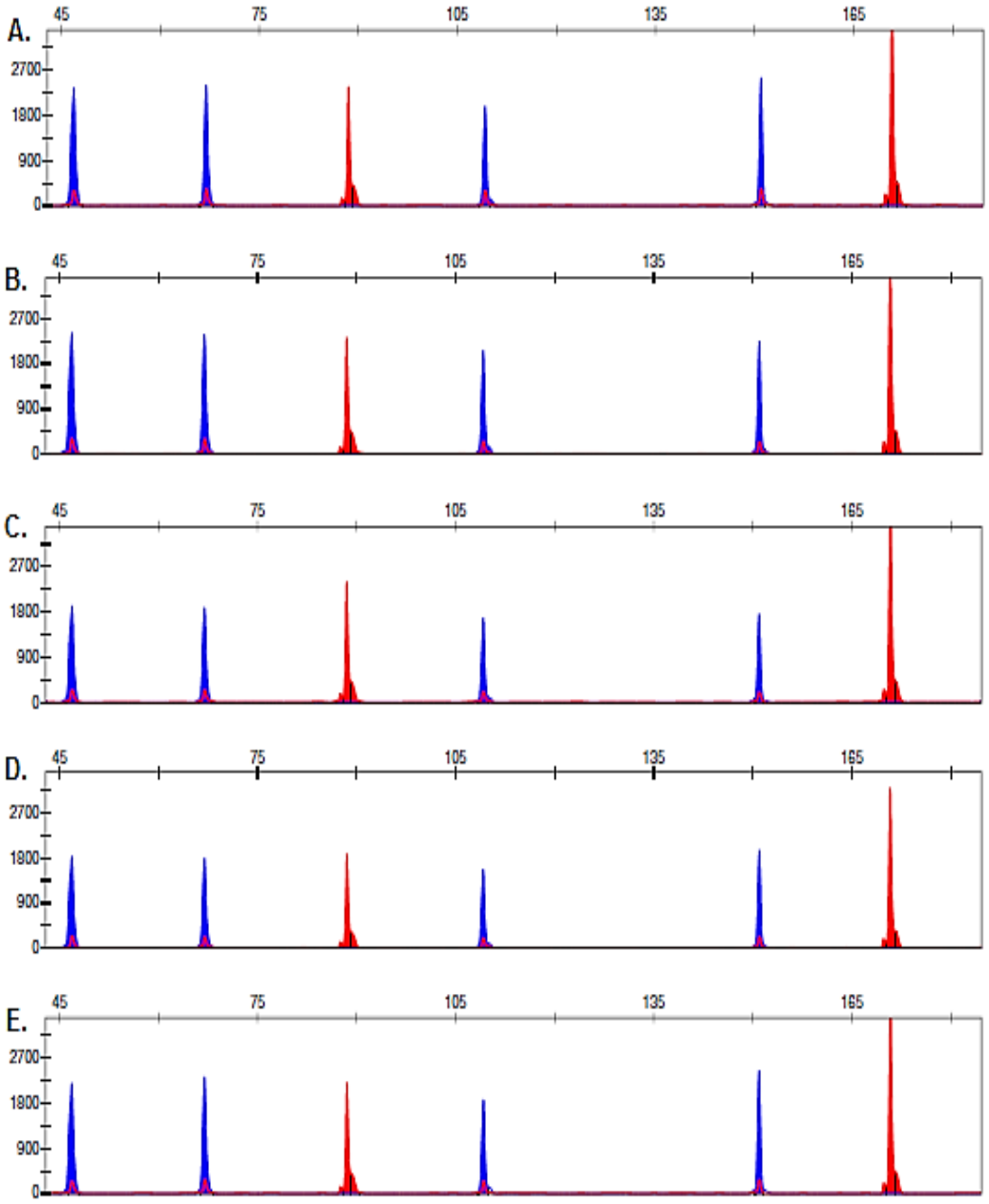


Figure 5-20: Electrophergrams generated from 1 ng of Human Genomic Female (Promega) control with final concentrations of (A) 30 ng/µl (B) 15 ng/µl, (C) 7.5 ng/µl, (D) 3.75 ng/µl and (E) 1.8 ng/µl CXR reference dye in 10 µl PCR final volume. The blue peaks are the mini 4-plex which to detect degradation, while the red peaks are the IACs, to detect inhibition.

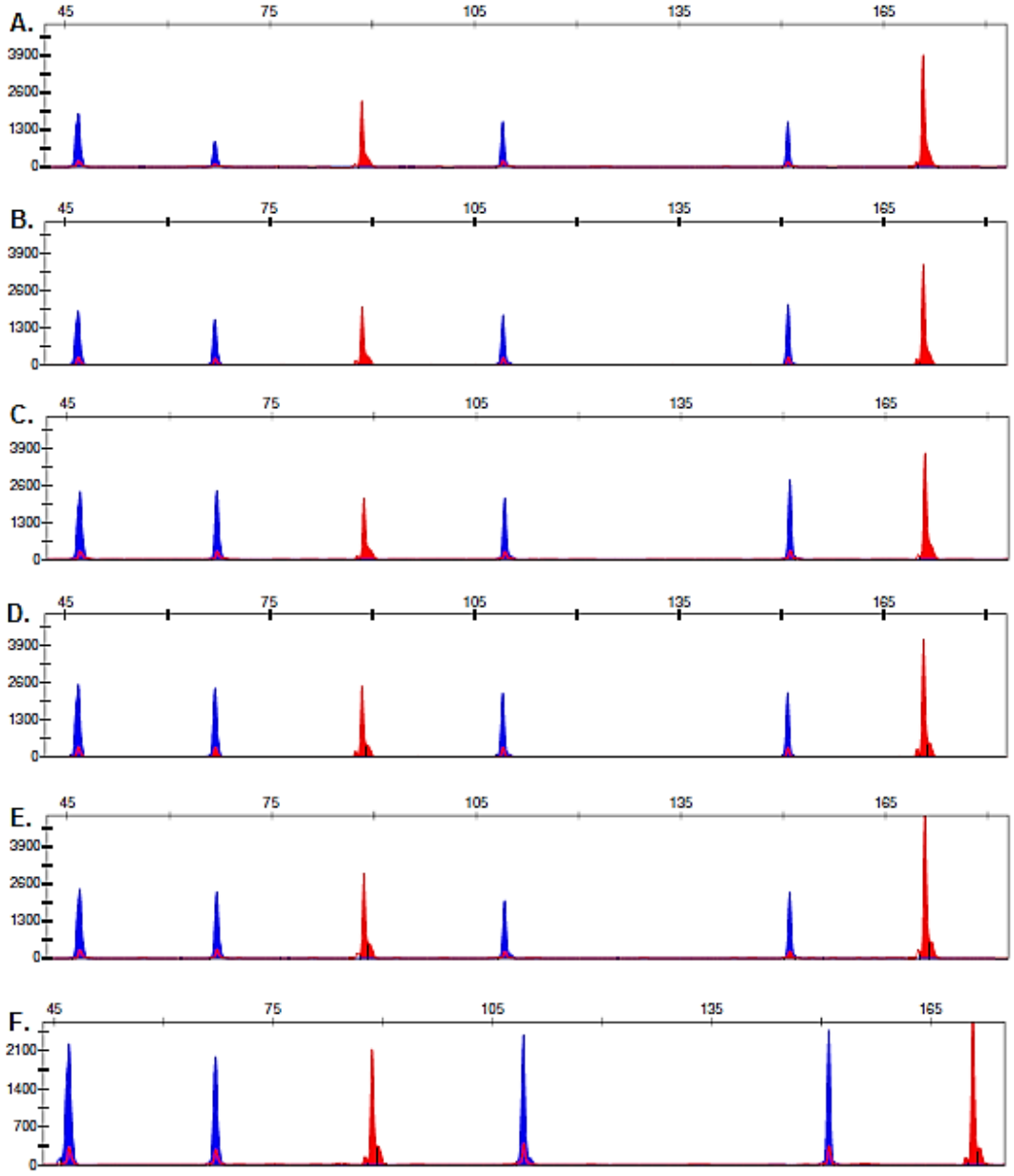


Figure 5-21**:** Electrophergrams generated from 1 ng of Human Genomic Female (Promega) control with final concentrations of (A) 1 M (B) 0.5 M, (C) 0.25 M, (D) 0.125 M ng and (E) 0.06 M DTT in 10 µl PCR final volume. (F) is the positive control with the concentration of 1 ng/µl. The blue peaks are the mini 4-plex which to detect degradation, while the red peaks are the IACs, to detect inhibition..

5.4 Discussion

Several studies have been described that commercial kits were used to assess the level of degradation (Hughes-Stamm, Ashton & Van Daal 2011, Dixon et al. 2006, Dixon et al. 2006). Most of the STR kits were fully utilized on the available information from DNA profiles obtained (e.g. the peak height), either from pristine or degraded DNA, and both at the allelic and genomic level. If degradation occur, the STR peaks often shifted, in relation to the size standard, presence of stochastic effect, allele drop out or missing all (Marasco et al. 2012). However, this method is not reliable for highly degraded DNA as it still could not differentiate null amplifications due to inhibition or simply due to insufficient quantities of template DNA.

In most laboratories, DNA degradation estimated by comparing the extent of amplification of a small fragment with a large fragment generated, simultaneously in the real time PCR assay. So do the PCR inhibition studies but mostly based on qualitative indication rather than quantitative measurement for PCR inhibitors (Seo et al. 2012). The PCR inhibition detected by an internal PCR control (IPC) during real time quantification. However, there are limitations with this approach as the amplicon in the real time PCR reactions is typically short, and does not necessarily reveal the full extent of PCR inhibition (Klein 2002). An alternative approach, which attempted to overcome these limitations, has to incorporate internal amplification controls (IACs) into the PCR reaction.

5.4.1 Multiplex design and validation

The purpose of developing the Mini 4-plex PCR assay is to understand the condition of degradation of the samples from previous research, which produced partial profiles with the 4-plex multiplex. Other than that, to assess the extent of DNA degradation and to assist in the choice of the informative DNA typing system available for highly degraded samples that has range between 50 bp to 150 bp (e.g. INDELs and SNPs). The Mini 4-plex that is specific to human and pig, was then combined with the modified two internal amplification controls IAC_{90} and IAC_{170}, to distinguish DNA degradation from PCR inhibitors (Nazir et al. 2013).

At the beginning stage, the Mini 4-plex primers designed from RAG-1. Nevertheless, problem arise when nonspecific peaks produced at 374 bp and 384 bp after the primers combined. Each primer pairs amplified with each other to investigate further the nonspecific peaks. This is due to the primers of Mini 4-plex were closed to each other;

therefore 70 bp forward amplified with the reverse of 112 bp and forward of 112 bp amplified with the reverse of 50 bp which formed the extra peaks. Since both RAG-1/2 co-expressed together, the 50 and 70 bp redesigned on RAG-2, and the other two amplicons 112 and 154 bp remained on RAG-1. When the new set of primers combined, no extra peaks were generated.

At first, the IACs used in this chapter were the ones designed in previous research; IAC_{70} and IAC_{410} (Zahra, Hadi & Goodwin 2012). However, there was a gap between the 154 bp amplicon of the mini 4-plex and the IAC_{410}. A modification in the reverse amplicon of IAC_{410} by reducing the target size to 170 bp. Therefore, the new Mini 4-plex & IACs is more susceptible to amplify highly degraded samples and assessed PCR inhibitors.

A range of temperature (56 - 62 °C) were studied and 60 °C was the optimum annealing temperature for the Mini 4-plex. Since the length of IAC primers were not significance different to the Mini 4-plex, the annealing temperature remained as 60 °C. The primer mix set four was chosen where the final working concentration were (0.06, 0.065, 0.06 and 0.065) µM for (50, 70, 112 and 154) bp respectively. All peaks balanced and no split peaks observed. The PCR master mix for 4-plex also remained the same even with the inclusion of IACs primers. Compared with the concentration used in primer mix of 4-plex, the Mini 4-plex is lower but still produced peak height more than 50 RFU.

5.4.2 DNA sensitivity study

After the development and optimisation of this multiplex, it was tested using different concentrations of template DNA to study its sensitivity. The results showed that this multiplex worked efficiently on DNA template as low as 0.009 ng, which highlighted the strength of the Mini 4-plex PCR assay to amplify small amounts of DNA. The optimum concentration for the Mini 4-plex is 0.6 ng as the peaks are balanced. Stochastic effects observed at 0.3 ng until 0.009 ng but no profile was obtained at negative control.

The sensitivity studies also showed that 50 bp and 70 bp amplicons persisted even the DNA template decreased, and peak height above 50 RFU. However, the larger amplicons, 112 bp and 154 bp were found to be sensitive to DNA degradation and presented reduction in peak height compared to the control. Alternative marker system that could utilize smaller amplicons such as INDELs and SNPs should be considered, where the amplicons could be small as 50 bp.

In comparison, the 4-plex system gave full profiles down to 0.07 ng. While at 0.03 ng and 0.01 ng DNA, weak and inconclusive results were obtained. (Iyavoo 2014) reported that the lowest concentration of the 4-plex to produce full profile was 0.1 ng and, partial profiles were developed until 0.05 ng but no profile below than that concentration.

One interesting observation from the sensitivity experiments was PowerPlex® 16 loci (Promega) tended to drop out first with decreasing of DNA concentrations. Full DNA profile were developed with 0.3 ng. The allele started to dropout at larger loci, which gave partial profiles at 0.15 ng and 0.06 ng. No DNA profile obtained below that concentration. According to (Tereba et al. 2002) and (McLaren 2007), PowerPlex® 16 System (Promega) gave full DNA profiles with minimum to 0.06 ng.

However, the peak heights range observed with the 4-plex system was 30 – 4800 RFU, whereas that with Mini 4-plex PCR assay the peak heights in the range of 60 –5200 RFU and PowerPlex® 16 was 52–2000 RFU.

The comparative study indicated that in-house 4-plex systems and the PowerPlex® 16 System (Promega) have low sensitivity level to detect highly degraded DNA. In addition, both of the systems have larger amplicons; 4-plex (70 bp - 384 bp) and PowerPlex® 16 (up to 400 bp), compared to the Mini 4-plex PCR assay.

5.4.2.1 Specificity

Complete profiles were developed for human and pig samples. This shows, the Mini 4-plex assay is specific to human and pig. Amplification products produced for sheep, roe and dog samples. However, the peaks are incomplete and imbalance. No profile was developed for rabbit and rat samples. In addition, there were no extra peaks observed in the Mini 4-plex & IACs system, only amplified the targeted loci.

5.4.3 DNA degradation study

This study is to assess the capacity of the Mini 4-plex PCR assay, to quantify DNA in highly degraded samples and to allow the visualization of the fragments which falls in the range of 50 to 154 bp, as the highly degraded DNA are more likely to be amplified by reducing the length of the amplicons. A 10-point degradation series was prepared by treating aliquots of high molecular weight Human Genomic DNA (Promega) with DNase I for increasing periods of time ranging from 2 to 180 mins (refer section 2.3.4). The results were a degradation series exhibiting incremental increase in the extent of DNA

degradation. Complete Mini 4-plex profiles were obtained reproducibly until 45 mins of DNase I digestion. The peak heights decreased as the length of degradation increased. At 60, 90 and 120 mins digested time, the larger amplicons; 112 and 154 bp decreased below 50 RFU but still can been seen. The 50 and 70 bp were callable until 120 mins of digestion. No profile was developed at 180 mins.

On comparison of the degradation study carried out by (Nazir 2012), complete 4-plex profiles developed up to 30 mins and above than 45 mins, no profile was obtained.

5.4.4 PCR inhibitor study

In this study, five potent PCR inhibitors; tannic acid, ethanol, TE buffer, phenol, CRX dye and DTT, were evaluated to test the efficiency of IACs in detecting inhibitors. All of these compounds can be found either in evidentiary items submitted in forensic laboratories or during DNA testing that can affect amplification of DNA samples.

As indicated in the results, if the IAC markers are reduced together with the Mini 4-plex at higher concentration, indicated the presence of inhibitors causing either partial or full inhibition. This can be seen in the peak height ratio of the IACs also the Mini 4-plex products.

As commonly used in the forensic community, the ratio of 0.6 for heterozygote balance was fixed to identify good DNA profiles; if the ratio of IAC peak height is above 0.6, that profile will be accepted (Bright, Turkington & Buckleton 2010, Petricevic et al. 2010, Kelly et al. 2012). Indirectly, it will also have indicated the performance of the PCR and quality of the DNA (Zahra, Hadi & Goodwin 2012). The heterozygote imbalance calculated by dividing the peak height of IAC_{170} with the peak height of IAC_{90}. Based on this ratio, the maximum PCR inhibitor concentration that could be present in the samples before this multiplex system failed to produce acceptable DNA profile was at 6.25 X TE Buffer; 25% and 6.25 % phenol; 30, 15, 3.75 and 1.8 ng/µl CXR dye and 1.0, 0.5, 0.25 and 0.06M DTT in a PCR reaction. This showed that was a decreased in the amplification efficiency due to partial inhibition.

Zahra et al 2011 reported that IAC_{90} was a robust marker and will drop out in the presence of high inhibitor concentration only after all human DNA failed to amplify, i.e. 100 and 50 ng/µl tannic acid, 9.6 and 4.8 % ethanol, 100X TE buffer and 99 and 50 % phenol. Similar to IAC_{410}, IAC_{170} found sensitive to PCR inhibitors and exhibited drop out at relatively higher concentrations (see Table 5-10 and 5-11).

Tannic acid is one of the common PCR inhibitors found. It is agent from leather (Opel, Chung & McCord 2010) and plant material (9) (Wilson 1997). It inhibits polymerase activities [(Schrader et al. 2012, Schrader et al. 2012). An earlier study of inhibitory effects on real-time PCR showed that reliable results could be obtained when tannic acid was less than 0.4 ng in a 25 µl PCR reaction (Kontanis, Reed 2006) The targeted fragment size used in that study was 132 bp Compared with the Mini 4-plex & IACs multiplex, reliable results could be observed at 25 ng and below, in 10 µl PCR reaction. However, for 4-plex (4-plex& IACs) system, the maximum amount of tannic acid before PCR inhibition detected was 50 ng in a 10 µl PCR reaction. This demonstrates that the new Mini-multiplex system is reliable and with the real-time PCR shows high level of PCR inhibitors, satisfactory profiles still can be obtaining with the multiplex.

Phenol-chloroform extraction is a liquid-liquid extraction method that separates mixtures of molecules based on the differential solubilities of the individual molecules in two different immiscible liquids. The excess of the phenol can degrade DNA polymerase (Bessetti 2007). Another study, showed that usage of phenol-saturated phosphate-buffered saline with concentrations as high as 15% did not affect the Tth DNA polymerase mediated reverse transcriptase activity (Katcher, Schwartz 1994). Whereas, the amplification of Mini 4-plex & IACs with Platinum® Multiplex PCR Master Mix (Thermo Fisher Scientific), the concentration of phenol as high as 25% did not inhibit the PCR, and using 4-plex & IACs system is up to 12.5 %. This showed that the Mini 4-plex & IACs system are more robust in detecting PCR inhibitors.

Other important sources of inhibitors are the materials and reagents that exposed to biological samples during processing or DNA purification, for example ethanol, EDTA in TE Buffer and DTT. These substances are used for DNA purification, DNA preservation and for cell lysis respectively, but they cause PCR inhibition only at relatively high concentrations (Schrader et al. 2012).

The maximum concentration of ethanol and TE buffer using the Mini 4-plex & IACs system are 3.5 % and 25 X respectively. In contrast, the 4-plex gave full profiles up to 5 % of ethanol and 1.25 X of TE buffer. Lastly, the CXR reference dye and DTT showed no inhibition in the amplified samples. This is maybe due to the concentrations are low and does not affect the amplification. Generally, the Mini 4-plex & IACs system is a sensitive multiplex to different inhibitors and reliable to assess DNA degradation besides observe PCR inhibition

5.5 Conclusion

This internal validation study provided laboratory experience and enabled the determination of the reliability and limitations of the new multiplex. The highlight of this study is this multiplex offers advantages over the commercial kits in that it is much less expensive also amplifying homozygous and heterozygous loci. The addition of the IACs allows the detection of PCR inhibition and differentiate between DNA degradation and PCR inhibition, thereby enhancing the multiplex's ability to be used as tool to assess degradation in different types of forensic evidence and to be used as a screening tool for complex forensic samples (Nazir et al. 2013). Moreover, the simplicity of the procedures described in this chapter should have applicability to a broad range of basic research and forensic applications. This assay is needed in forensic laboratory especially with laboratory with a small budget. In addition, a good sample management could be applied as no samples or kits would be into waste.

CHAPTER 6

APPLICATION OF NEW MINI 4-PLEX SYSTEM (MINI 4-PLEX & IACS)

In the first part of this chapter was undertaken by two members of the Forensic Genetics Group, UCLAN which includes myself and MRes student, Balnd Mustafa Albarzinji, from April 2014 to September 2015.

6.1 Overview

The overview of real-time PCR technology has significantly improved and simplified the quantification of nucleic acids, and this technology has become an invaluable tool for many scientists working in different disciplines (Klein 2002). Even though, the use of real-time PCR methods has widespread, it also highlighted some of the critical points and limitations of these assays. This includes, during DNA quantification using real-time PCR, the actual quality of the samples cannot be determined and this can cause difficulty to obtain good DNA profiles.

In the previous chapter it was demonstrated that the new Mini 4-plex multiplex (Mini 4-plex & IACs) was effective at assessing DNA degradation and PCR inhibition. Further experiments were designed to evaluate whether the multiplex could be useful for quantifying the amount of DNA present using the peak heights of the electropherograms; this method was compared to an in-house and a commercial real-time quantification method.

Therefore, Chapter 6 describes efforts to use the Mini 4-plex assay to determine the concentration of DNA in a sample.

6.1.1 Objectives

- To study the behaviour of the New Mini 4-plex system using extracted DNA from saliva samples that simulated to be recovered from the crime scene by distilled water and cell lysis buffer at various temperatures and times.
- To study the combined Mini 4-plex and IACs multiplex PCR and to assess the capability of the internal amplification control markers, 90 bp and 170 bp, to detect

the presence of inhibitors using extracted DNA from pig bones samples (femur and rib) that had been exposed to the environment.

- To compare the accuracy and precision of the multiplex for DNA quantification using Mini 4-plex and commercial real time system on swabbed samples recovered from real crime scene by distilled water and cell lysis buffer.

6.2 Materials and methods

6.2.1 Human DNA saliva samples

Extracted DNA from saliva samples (these samples were provided by Dinah Bandar N Aloraer, PhD student, UCLAN) were used to test the mini 4-plex multiplex PCR. The saliva samples were stored at different temperatures (-20 °C, room temperature (RT), 37 °C and 50 °C) for a range time of (3 and 48 hr). The samples were recovered using two different buffer, distilled water and cell lysis, to compare the results of the recovery.

6.2.2 Pig bone samples

Extracted DNA from pig bone samples (femur and rib) were used to study the combined mini 4-plex & IACs multiplex PCR and to assess the capability of the IAC_{90} and IAC_{170} to detect the presence of inhibitors (these samples were provided by Ali Abduljaleel Al-Janabi, PhD student, UCLAN). The bone samples had been exposed to the environment; femurs exposed for 1 and 4 years respectively and ribs for 1, 3 and 4 years respectively. Then the bones were extracted using various extraction methods which were suitable to extract degraded samples. Three extraction methods; PrepFiler® Forensic DNA Extraction Kit (Applied Biosystems), DNeasy Blood & Tissue Kit (Qiagen) and ChargeSwitch® gDNA Mini Tissue Kit (Thermo Fisher Scientific) were used to purify DNA from bone samples.

6.2.3 Comparing 4-plex and Mini 4-plex

Quantitative data was reanalysed and relationship between peak height at different experimental conditions were looked for.

6.3 Results

6.3.1 Application of Mini 4-plex & IAC multiplex system on simulated forensic samples.

6.3.1.1 Quantitation of saliva samples using Real Time PCR

The source of extracted saliva and bones samples were described at paragraph 6.2 above. Extracted DNA from saliva samples were quantified using the real-time PCR with the Human Quantifiler Kit (Applied Biosystems) according to manufacturer's instructions (refer Section 2.5.2.1). The mixture was prepared at a reduced final volume of 12.5 µl including: Quantifiler Human Primer Mix (5.25 µl), master mix (6.25 µl), DNA template (1 µl). Serial dilution of Human DNA Standard control (200 ng/µl) was prepared (ranging from 50 ng/µl to 0.023 ng/µl). Quantification was carried out using Applied Biosystems 7500 real-time PCR System with the following parameters: Stage 1 (1cycle) 95 °C for 10 min followed by Stage 2 (40 cycles) 95 °C for 15 s, 60 °C for 1 min.

It can be observed from Table 6-1 and 6-2 that the standard deviation of the quantification process of saliva samples using distilled water and lysis buffer in (3hr and 48 hr) at various temperature were relatively low. The results obtained show that there is not much difference between actual concentrations of the samples with the estimated concentration from real-time PCR. There was error observed in the Quantifiler® Human DNA Quantification kit (Applied Biosystems). In contrast, the R.S.D are high above 10% until the lowest concentration, 0.22 ng, meaning that this kit is not robust for the quantification of samples with low template of DNA.

Table 6-1**:** Table shows the quantitation results of extracted DNA from saliva samples and collected with distilled water after 3 hr and 48 hr at different temperatures, -20, Room Temperature (RT), 37 and 50 °C respectively, using Quantifiler Human DNA Quantification Kit (Applied Biosystems).

Hour	Temp (°C)	Sample 1 (ng/µl)	Sample 2 (ng/µl)	Sample 3 (ng/µl)	Avg (ng/µl)	s.d.	R.S.D (%)
3 hr	-20	2.33	1.15	0.62	1.36	0.87	63.97
	RT	1.22	1.59	0.75	1.18	0.42	35.59
	37	1.00	0.65	0.96	0.87	0.19	22.02
	50	0.77	0.40	0.46	0.54	0.19	35.1
48 hr	-20	12.52	8.37	11	10.63	2.09	19.66
	RT	0.78	2.16	3.38	2.10	1.3	61.9
	37	0.22	0.10	0.16	0.16	0.06	37.5
	50	0.30	0.13	0.34	0.25	0.1	40

Table 6-2**:** Table shows the quantitation results of extracted DNA from saliva samples and collected with cell lysis buffer after 3 hr and 48 hr at different temperatures, -20, Room Temperature (RT), 37 and 50 °C respectively, using Quantifiler Human DNA Quantification Kit (Applied Biosystems).

Hour	Temp (°C)	Sample 1 (ng/µl)	Sample 2 (ng/µl)	Sample 3 (ng/µl)	Avg (ng/µl)	s.d.	R.S.D (%)
3 hr	-20	0.92	1.24	2.53	1.56	0.84	53.84
	RT	4.00	5.29	1.93	3.74	1.69	45.18
	37	1.76	3.43	2.39	2.52	0.84	33.3
	50	1.87	0.62	0.75	1.08	0.68	62.96
48 hr	-20	5.17	2.05	6.04	4.42	2.09	47.2
	RT	4.29	3.77	2.69	3.58	0.81	22.62
	37	2.65	1.16	1.86	1.89	0.74	39.15
	50	2.65	3.25	3.19	3.03	0.33	10.91

6.3.1.2 DNA amplification of saliva samples using Mini 4-plex and 4-plex multiplex PCR

Amplification of the extracted DNA from saliva samples were carried out using both Mini 4-plex and 4-plex system respectively to study the stability of the markers and the variation in the height of the peaks between Mini 4-plex and 4-plex. 5 µl of Platinum® Multiplex PCR Master Mix was used and mixed with 1 µl of the extracted DNA, 0.6 µl primer mixture and 3.4 µl of nuclease-free water was added to adjust the reduced final volume of 10 µl. The electropherogram results of the Mini 4-plex and 4-plex multiplex PCR showed that there was degradation in the samples that were collected by distilled water especially after 48 hr collection at 37 °C and 50 °C. While the results for samples that were collected by cell lysis were more stable and balanced. However, DNA degradation can be observed with the samples of 48 hr post-collection (see Figure 6-4).

Table 6-3**:** The average (avg), standard deviation (s.d.) and relative standard deviation (R.S.D %) of the peak heights of Mini 4-plex multiplex PCR markers using extracted DNA from saliva samples collected by distilled water after 3 and 48 hr respectively.

Hour	Temp (°C)	Peak height (RFU)						
		Markers				Avg	s.d	R.S.D (%)
		50	70	112	154			
3 hr	-20	1640	1824	1311	1274	1512	265	17.5
	RT	1672	1964	1174	1293	1525	361.7	23.6
	37	1162	1454	607	674	974.2	404.3	41.5
	50	1600	1606	792	872	1217.5	446.3	36.6
48 hr	-20	2844	2938	2445	2491	2679.5	247.9	9.25
	RT	1375	1656	533	475	1009.7	595.6	58.9
	37	315	309	106	145	218.7	108.8	49.7
	50	247	243	93	102	171.2	82.2	49.7

Table 6-4**:** The average (avg), standard deviation (s.d.) and relative standard deviation (R.S.D %) of the peak heights of Mini 4-plex multiplex PCR markers using extracted DNA from saliva samples collected by lysis buffer after 3 and 48 hr respectively.

Hour	Temp (°C)	Peak height (RFU)						
		Markers				Avg	s.d	R.S.D (%)
		50	70	112	154			
3 hr	-20	2530	2305	2540	2179	2388.5	176.8	7.4
	RT	2403	2145	2392	2141	2270.2	147	6.4
	37	2388	2176	2432	2120	2279	154	6.7
	50	2207	2035	2182	1724	2037	222	10.9
48 hr	-20	3139	2964	3424	3315	3210.5	201.9	6.29
	RT	3424	3296	3238	3186	3286	102.3	3.12
	37	2686	2774	2114	2083	2414	366.5	15.1
	50	3389	3075	2866	2947	3069	229.8	7.5

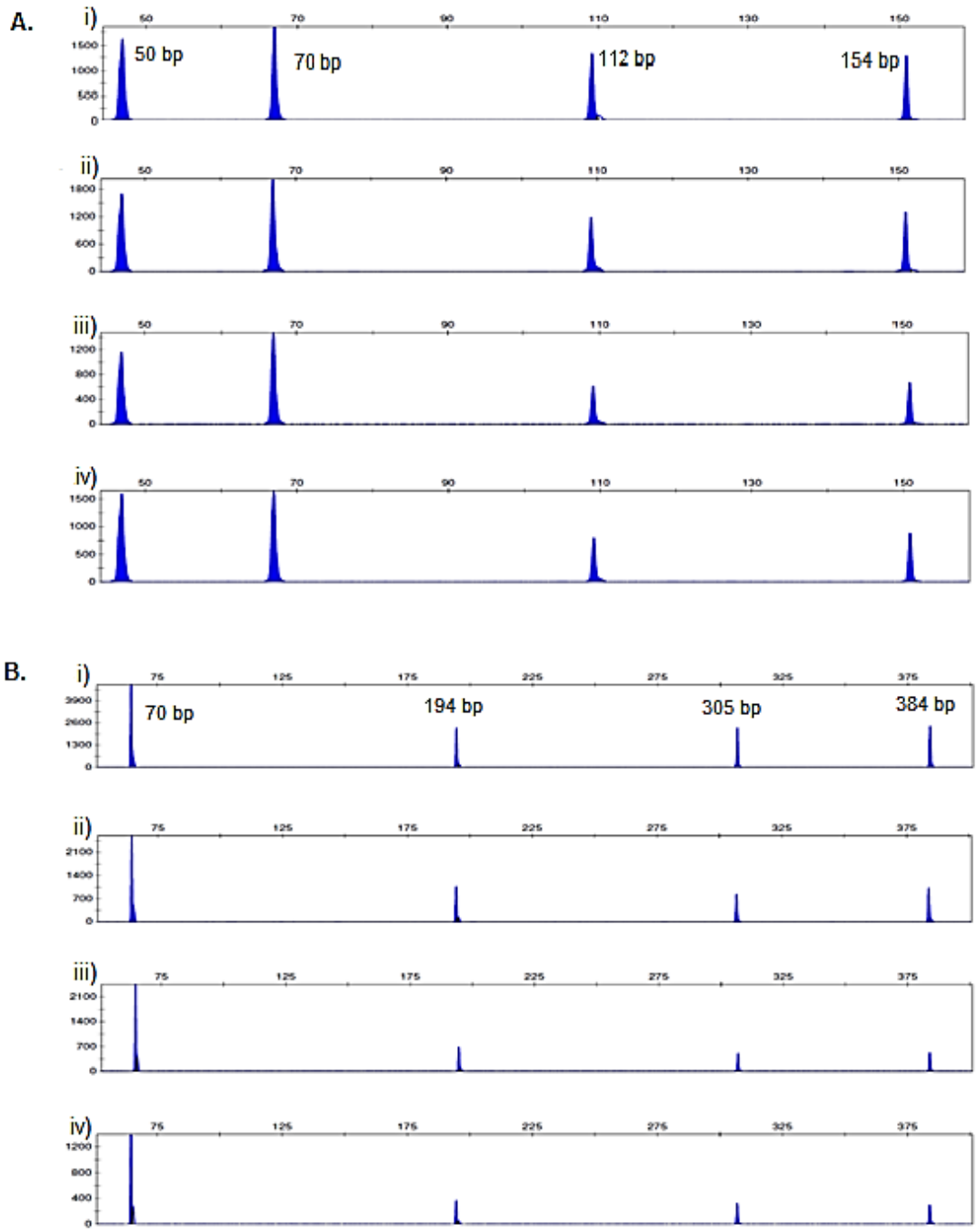


Figure 6-1: Electropherograms developed from (A) Mini 4-plex multiplex and (B) 4-plex PCR amplification of extracted DNA from saliva samples after 3 hr collection by distilled water at different temperatures (i) 20 °C (2.3 ng), (ii) RT (1.5 ng), (iii) 37 °C (1 ng) and (iv) 50 °C (0.7 ng).

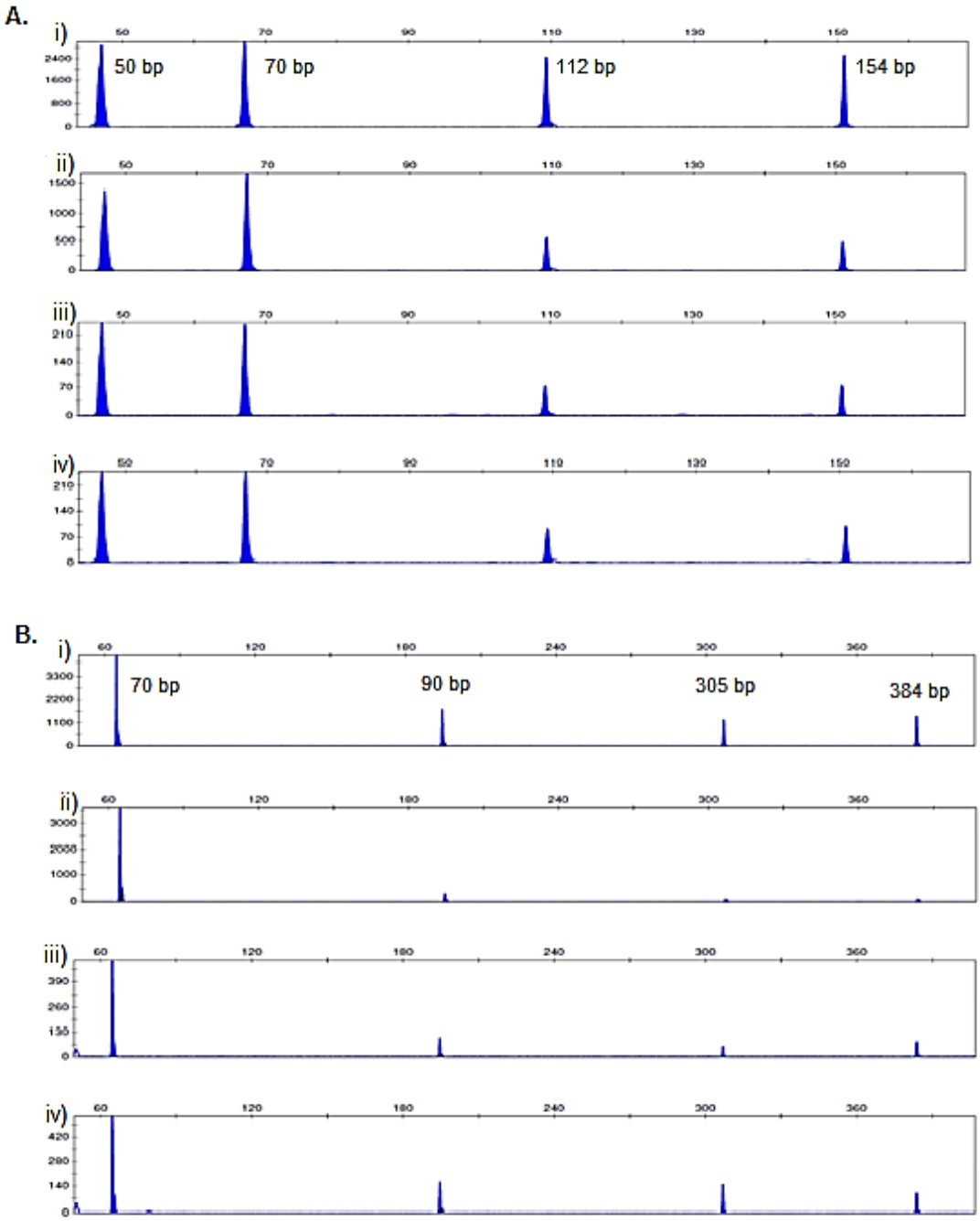


Figure 6-2**:** Electropherograms developed from (A) Mini 4-plex and (B) 4-plex multiplex PCR amplification of extracted DNA from saliva samples after 48 hr collection using distilled water at different temperatures (i) -20 °C (11 ng), (ii) RT (2.1 ng), (iii) 37 °C (0.22 ng) and (iv) 50 °C (0.13 ng).

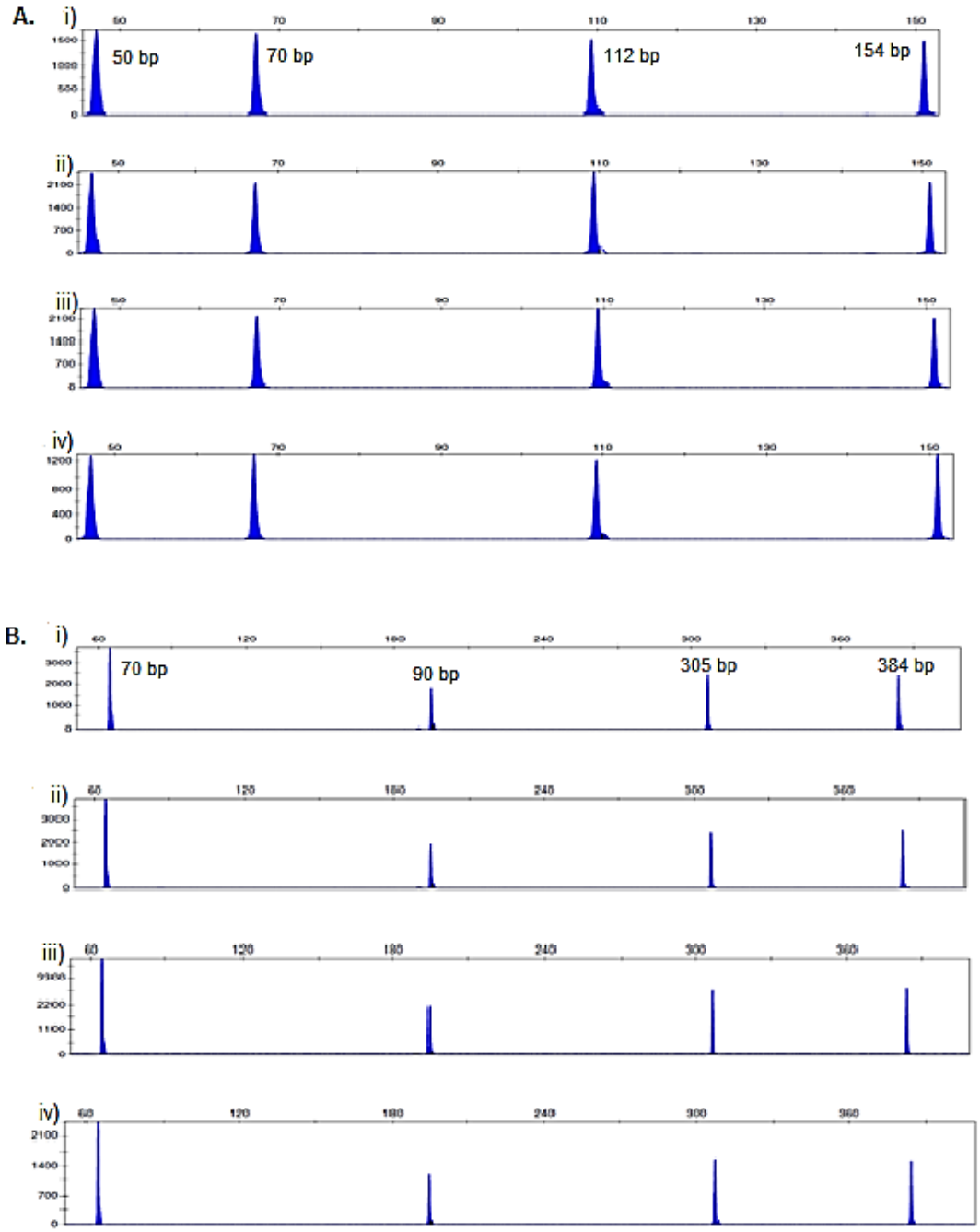


Figure 6-3: Electropherograms developed from (A) Mini 4-plex and (B) 4-plex multiplex PCR of saliva samples after 3 hr collection by cell lysis at different temperatures (i) -20 °C (1.24 ng), (ii) RT (1.93 ng), (iii) 37 °C (1.76 ng) and (iv) 50 °C (0.75 ng).

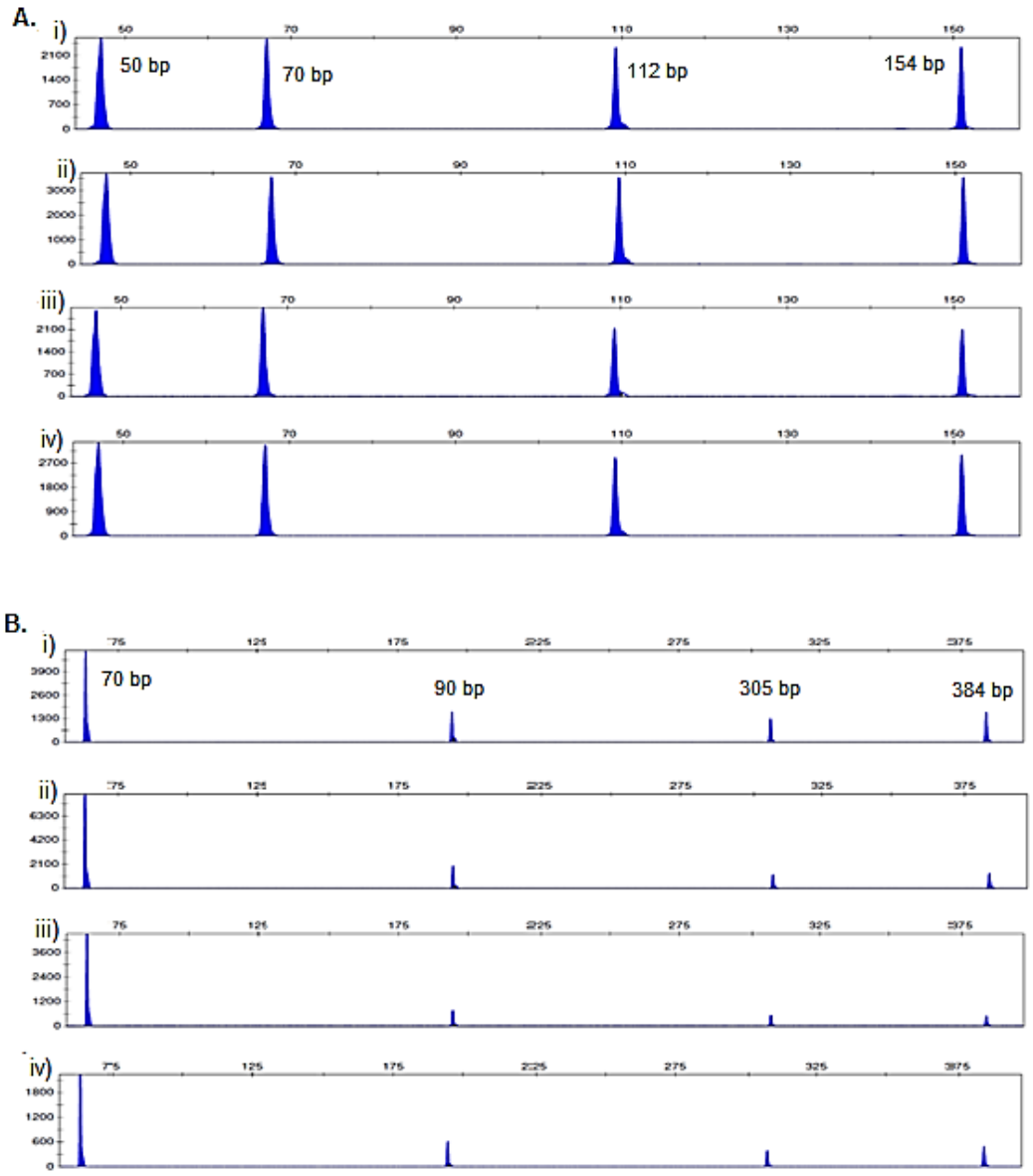


Figure 6-4**:** Electropherograms developed from (A) Mini 4-plex and (B) 4-plex multiplex PCR of extracted DNA saliva samples after 48 hr collection by cell lysis at different temperatures (i) -20 °C (2.05 ng), (ii) RT (3.77 ng), (iii) 37 °C (1.86 ng) and (iv) 50 °C (3.25 ng).

6.3.1.3 Statistical analysis of saliva samples amplified by mini 4-plex

Statistical Analysis of Variance (ANOVA) was carried out using R Studio software to study the F value and to see the differences of the peak height of the mini 4-plex multiplex PCR amplification of saliva samples that were collected at two different times, 3 hr and 48 hr using distilled water and cell lysis, separately. The ANOVA results show that there is no significant difference among groups of the peak heights of saliva samples collected by distilled water after 3 hr at different temperatures ($F(3, 12) = 1.976$, $p = 0.171$), but there is a significant difference among groups of peak heights of the samples collected after 48 hr using distilled water ($F(3, 12) = 50.43$, $p = 4.47e\text{-}07$). The ANOVA results for the saliva samples were collected by cell lysis also show that there is no significant difference among peak height groups of the samples collected after 3 hr ($F(3, 12) = 2.78$, $p = 0.0867$), but the difference is statistically significance among groups of the peak heights for the samples collected after 48 hr using cell lysis buffer at different temperatures ($F(3, 12) = 10.6$, $p = 0.001$) as the p-value is less than 0.05, the significance level.

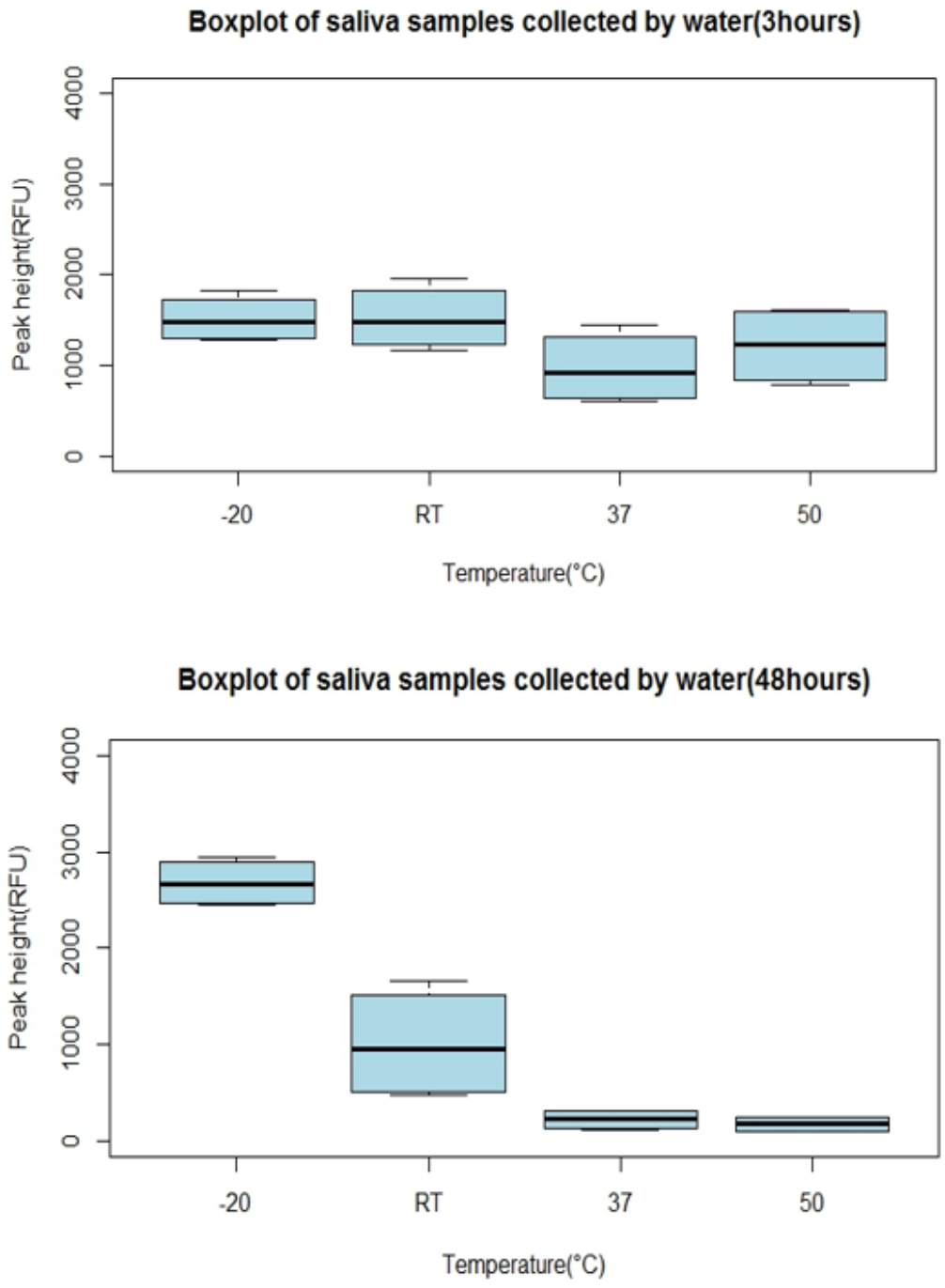


Figure 6-5**:** Boxplots generated from height of peaks of saliva samples collected after 3 and 48 hr using distilled water at four different temperatures (-20 °C, Room Temperature (RT), 37 °C and 50 °C), amplified using mini 4-plex multiplex PCR.

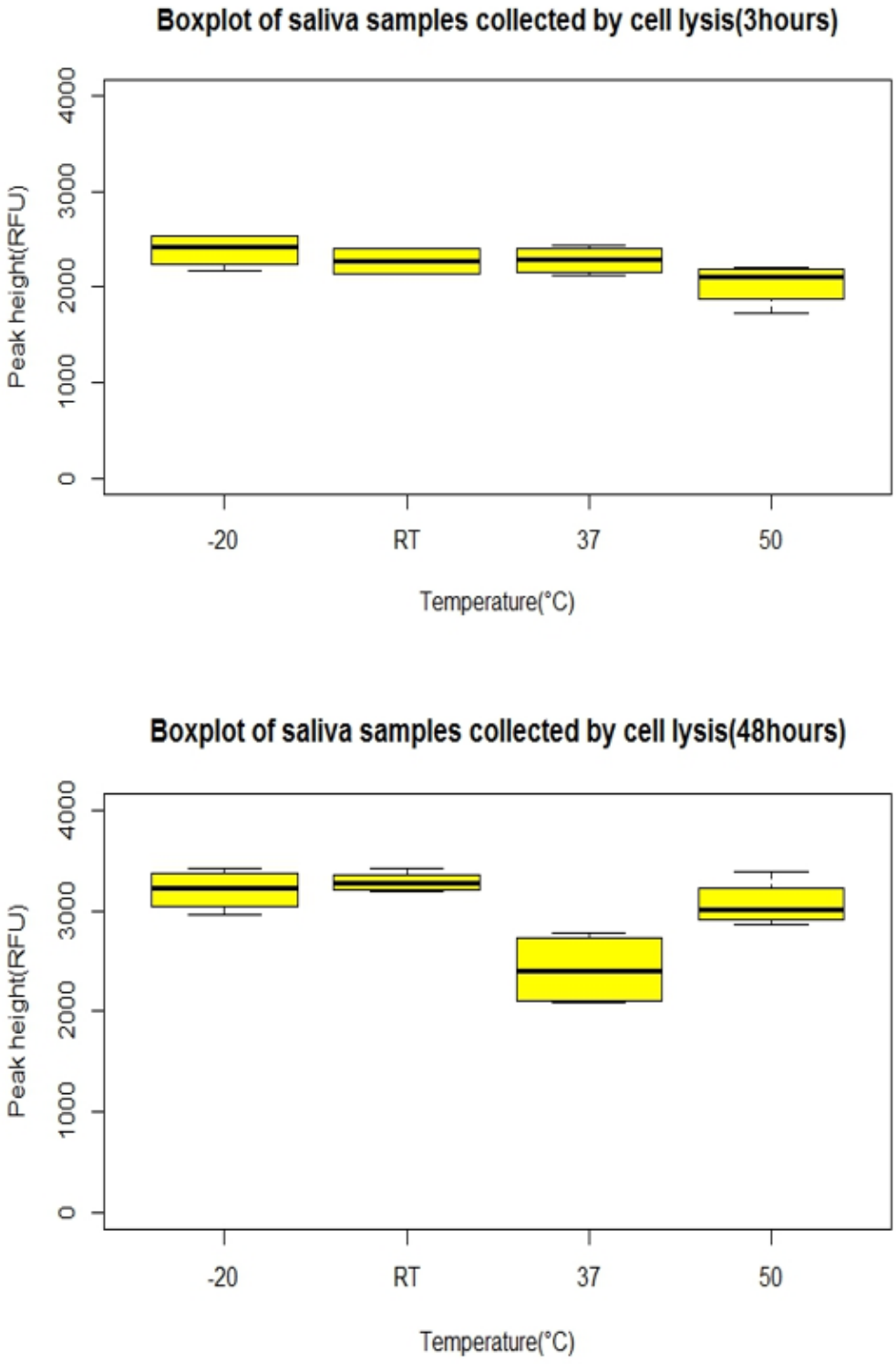


Figure 6-6**:** Boxplots generated from height of peaks of saliva samples collected by cell lysis after 3 and 48 hr at four different temperatures (-20 °C, Room Temperature (RT), 37 °C and 50 °C), amplified using mini 4-plex multiplex PCR.

6.3.1.4 Amplification of pig samples using mini 4-plex and IACs multiplex PCR

Extracted DNA from pig bone samples (femur and rib) were used to study the combined mini 4-plex and IACs multiplex PCR and to assess the capability of the IAC90 and IAC170 to detect the presence of inhibitors. The bones had been exposed to the environment (femur 1 and 4 years, rib 1, 3 and 4 years) to study the extraction methods on degraded DNA. Three extraction methods (PrepFiler® Forensic DNA Extraction Kit (Applied Biosystems), DNeasy Blood & Tissue Kit (Qiagen) and ChargeSwitch® gDNA Mini Tissue Kit (Thermo Fisher Scientific) were used to purify DNA from bone samples. Amplification of the extracted DNA was carried out using Platinum® Multiplex PCR Master Mix (5 µl) and mixed with 1 µl of the extracted DNA, 1 µl of IACs template, 0.6 µl of primer mixture and 2.4 µl of nuclease-free water was added to adjust the reduced final volume of 10 µl. The amplification was performed using Veriti® 96-Well Thermal Cycler (Thermo Fisher Scientific) and the PCR products were run on ABI 3500 Genetic Analyzer.

The electrophergram shows that DNA samples extracted using both methods PrepFiler® Forensic DNA Extraction Kit and DNeasy Blood & Tissue Kit produced higher peak heights of the larger amplicon IACs 170 bp, with the ratio above 0.6 by dividing the peak height of IAC_{170} with the peak height of IAC_{90}. While inhibition was detected in the extracted rib samples using ChargeSwitch® gDNA Mini Tissue Kit as lower peak heights of the larger marker was generated and the ratio of the peak heights were below 0.6 (rib 3 year = 0.54) and (rib 4 year = 0.56). Amplification of the mini 4-plex markers were obtained in the extracted femur samples using PrepFiler® Forensic DNA Extraction Kit and the two shorter amplicons, 50 bp and 70 bp, were amplified in the rib 1-year sample with RFU above 50.

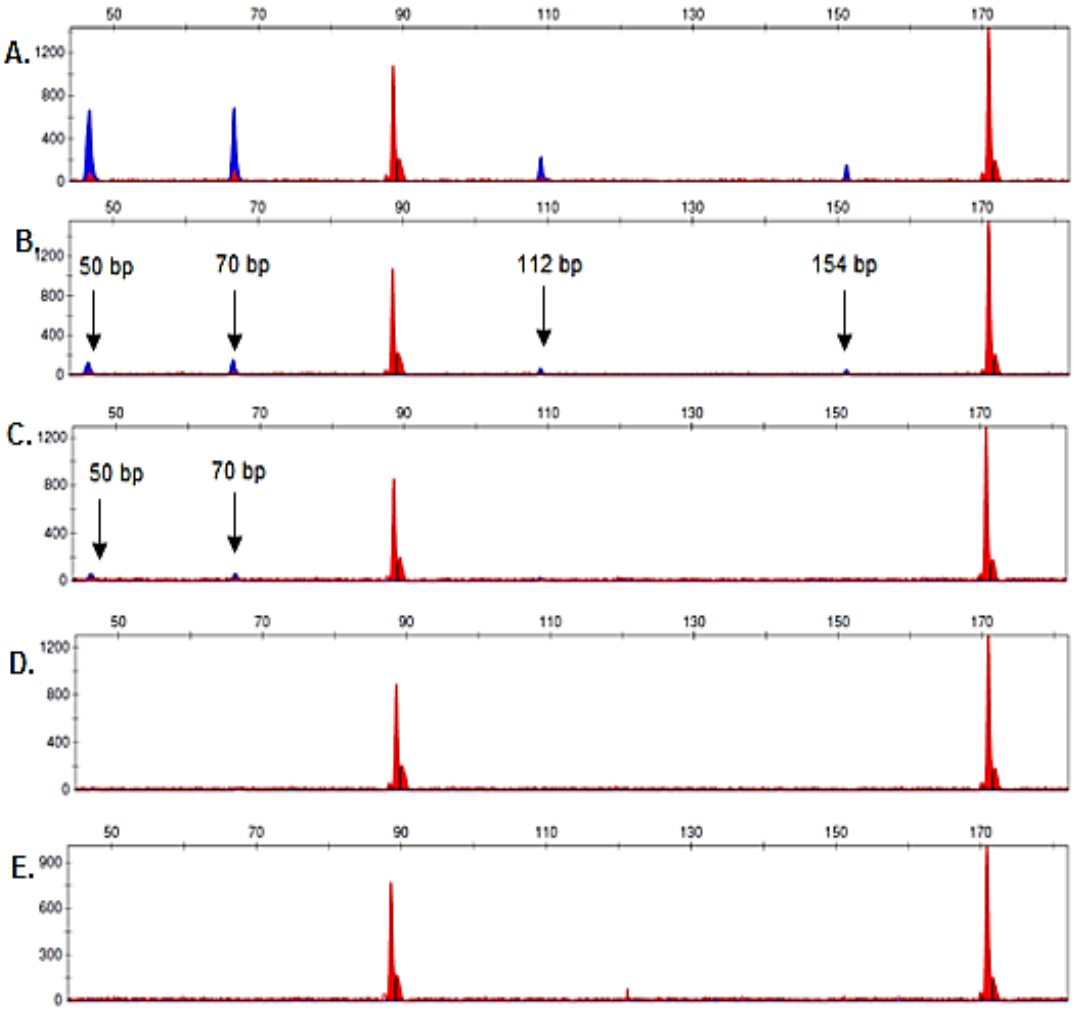


Figure 6-7: Electropherograms shows amplification of the extracted bone samples using PrepFiler® Forensic DNA Extraction Kit for (A) Femur 1 year, (B) Femur 4 year, (C) Rib 1 year, (D) Rib 3 year and (E) Rib 4 year. The blue peaks are the mini 4-plex which to detect degradation, while the red peaks are the IACs, to detect inhibition.

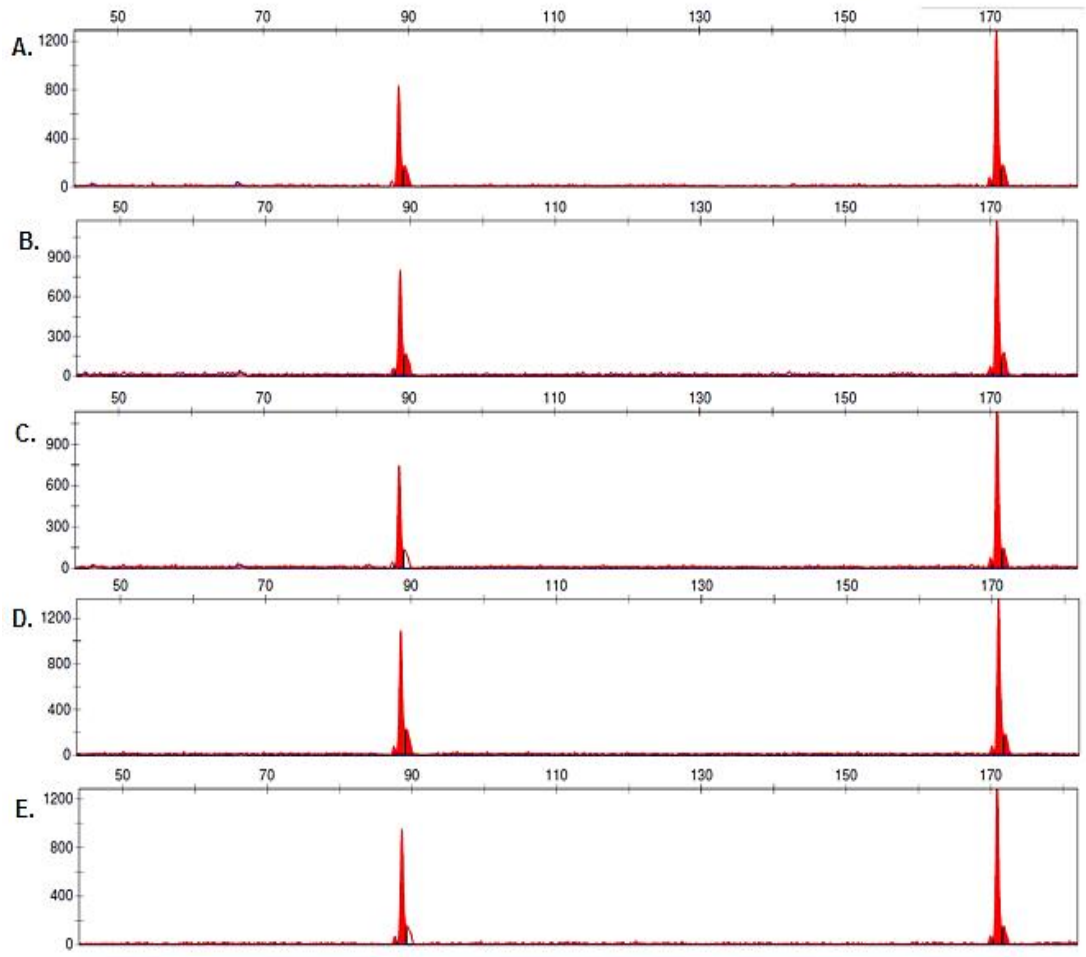


Figure 6-8: Electropherograms shows amplification of the extracted bone samples using DNeasy Blood & Tissue Kit for (A) Femur 1 year, (B) Femur 4 year, (C) Rib 1 year, (D) Rib 3 year and (E) Rib 4 year. The red dyed alleles are the IACs, to detect inhibition.

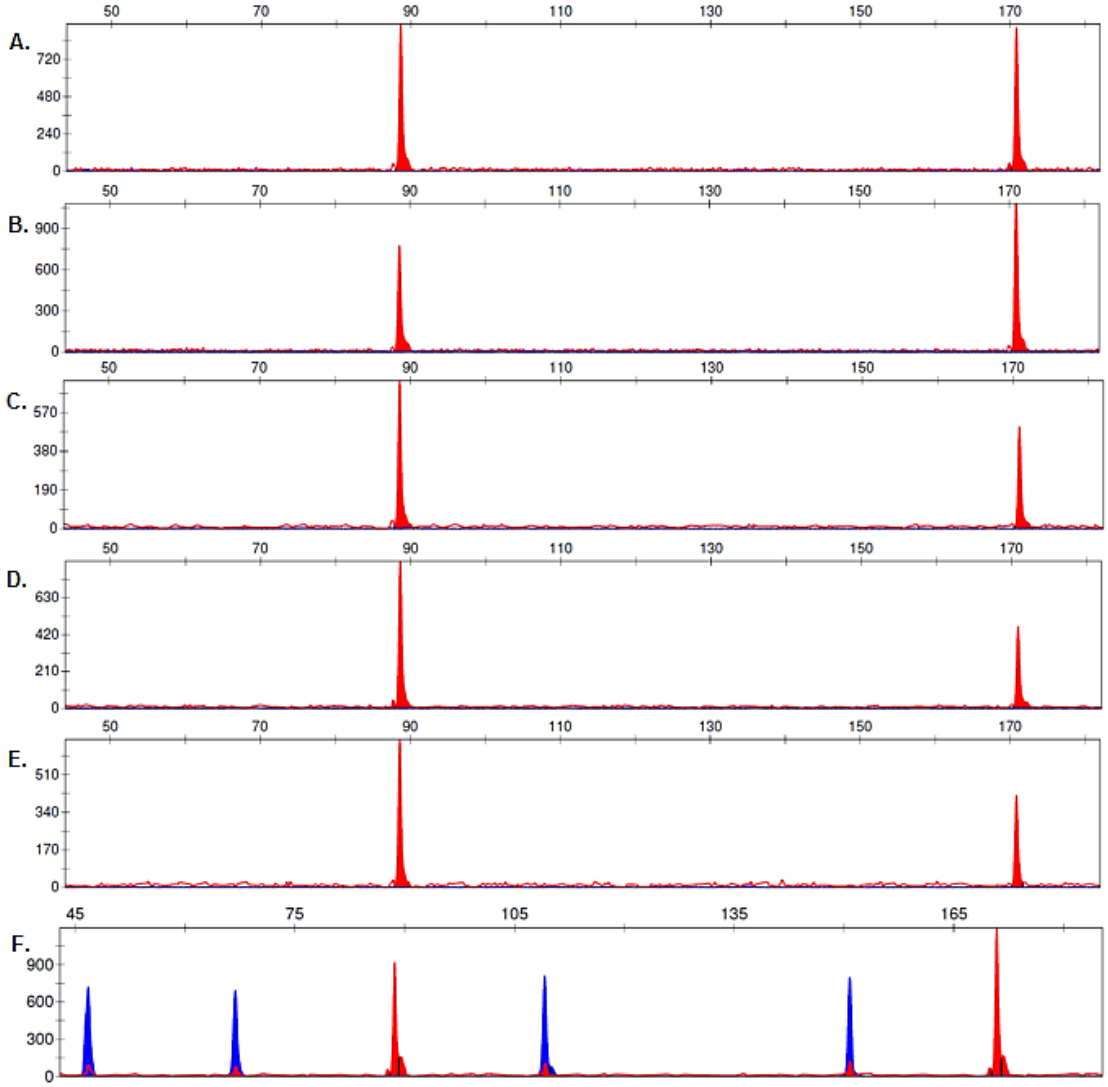


Figure 6-9**:** Electropherograms shows amplification of the extracted bone samples using ChargeSwitch® gDNA Mini Tissue Kit for (A) Femur 1 year, (B) Femur 4 year, (C) Rib 1 year, (D) Rib 3 year, (E) Rib 4 year and (F) positive control. The blue peaks are the mini 4-plex which to detect degradation, while the red peaks are the IACs, to detect inhibition..

Table 6-5: Amplification of the extracted bone samples (femur and rib) using mini 4-plex & IACs multiplex PCR.

Method	Sample	50 bp	70 bp	112 bp	154 bp	IAC $_{90}$	IAC $_{170}$
PrepFiler®	Femur 1-year	+++	+++	++-	++-	+++	+++
PrepFiler®	Femur 4-year	++-	++-	+--	+--	+++	+++
PrepFiler®	Rib 1-year	+--	+--	---	---	+++	+++
PrepFiler®	Rib 3-year	---	---	---	---	+++	+++
PrepFiler®	Rib 4-year	---	---	---	---	+++	+++
DNeasy	Femur 1-year	---	---	---	---	+++	+++
DNeasy	Femur 4-year	---	---	---	---	+++	+++
DNeasy	Rib 1-year	---	---	---	---	+++	+++
DNeasy	Rib 3-year	---	---	---	---	+++	+++
DNeasy	Rib 4-year	---	---	---	---	+++	+++
ChargeSwitch®	Femur 1-year	---	---	---	---	+++	+++
ChargeSwitch®	Femur 4-year	---	---	---	---	++-	+++
ChargeSwitch®	Rib 1-year	---	---	---	---	+++	++-
ChargeSwitch®	Rib 3-year	---	---	---	---	+++	++-
ChargeSwitch®	Rib 4-year	---	---	---	---	+++	++-

Note: “+++” represents the presence and “---” represents the absence of the amplicons of mini 4-plex & IACs.

6.3.2 Quantification using the multiplex (Mini 4-plex & IACs)

6.3.2.1 Quantification of serial dilution samples using real-time PCR

A serial dilution was carried out using stock 193 ng/µl Human Genome Female Control DNA (Promega) until the final concentration was 0.05 ng/µl. Each point of the concentration was prepared in triplicate and the final volume of each samples was 10 µl. This serial dilution samples were quantified using Quantifiler® Human DNA Quantification kit. The quantification results are shown in Table 6-6.

Table 6-6: Table below shows the concentrations of the DNA samples which were prepared by serial dilution using control DNA 9947A and quantified using Quantifiler® Human DNA Quantification kit (Applied Biosystems).

DNA concentration ng/µl	Quantification using Quantifiler® Human DNA Quantification kit (ng/µl)					
	Set A	Set B	Set C	Avg.	s.d.	R.S.D (%)
100	104.2	98.7	102.3	101.733	2.793	2.746
50	51	49.7	47.7	49.467	1.662	3.361
25	25.3	26	25.7	25.667	0.351	1.368
10	10.5	11.1	10.4	10.667	0.379	3.549
5	5.2	4.8	5.1	5.033	0.208	4.136
2.5	2.48	2.47	2.5	2.483	0.015	0.615
1.25	1.265	1.3	1.324	1.296	0.030	2.289
0.6	0.606	0.572	0.583	0.587	0.017	2.956
0.3	0.31	0.293	0.322	0.308	0.015	4.726
0.15	0.105	0.098	0.103	0.102	0.004	3.535
0.07	0.07	0.071	0.068	0.070	0.002	2.193
0.03	0.027	0.029	0.028	0.029	0.001	3.509
0.01	0.013	0.012	0.013	0.013	0.001	4.558

Note: Avg.: Average, s.d,: Standard deviation, R.S.D: Relative Standard Deviation

The results obtained show that there is not much difference between actual concentrations of the control DNA with the estimated concentration from real-time PCR. There was no error observed in the Quantifiler® Human DNA Quantification kit (Applied Biosystems) where the R.S.D stayed below 5% until the lowest concentration, 0.01 ng. Thus, meaning that this kit is more robust for the quantification of samples with low template of DNA.

6.3.2.2 Amplification and analysis of the serial dilution using the Mini-4plex & IACs system.

After the serial dilution samples were quantified using real-time PCR, the remaining samples were used to amplify the Mini-4plex & IACs system using 1 µl of each samples and the results obtained from the electropherogram was shown in Table 6-7.

Table 6-7: Table below shows the average peak heights of the electropherograms produced using the serial diluted Human Genome Control DNA (Promega) samples.

DNA concentration ng/µl	Average peak height (RFU)					
	Set A	Set B	Set C	Avg.	s.d.	R.S.D (%)
100	2783.0	2810.0	2701.0	2764.7	56.766	2.053
50	2583.5	2346.1	2734.0	2554.5	195.566	7.656
25	3477.5	3500.0	3254.6	3410.7	135.654	3.977
10	3162.3	3014.2	3545.8	3240.8	274.356	8.466
5	3405.5	3510.2	3554.6	3490.1	76.555	2.193
2.5	1825.8	1789.2	1897.3	1837.4	54.986	2.993
1.25	1199.8	1026.7	1206.1	1144.2	101.793	8.897
0.6	695.50	685.20	690.00	690.20	5.1540	0.747
0.3	201.00	210.00	199.30	203.40	5.750	2.827
0.15	189.00	187.60	173.25	183.30	8.7170	4.756
0.07	72.500	66.25	89.50	76.100	12.032	15.814
0.03	0.00	0.00	0.00	0.00	0.00	0.00
0.01	0.00	0.00	0.00	0.00	0.00	0.00

Note: Avg.: Average, s.d.: Standard deviation, R.S.D.: Relative Standard Deviation

The results showed that as the concentration decreased, precision also decreased. This can be seen at the concentration of 0.03 ng/ul where the R.S.D decrease to 0%. This is mainly because of the dropout alleles at the Mini 4-plex were generated at those concentrations. Full profiles were obtained from 0.07 ng/ul and above. The electropherograms developed is shown in Figure 6-10.

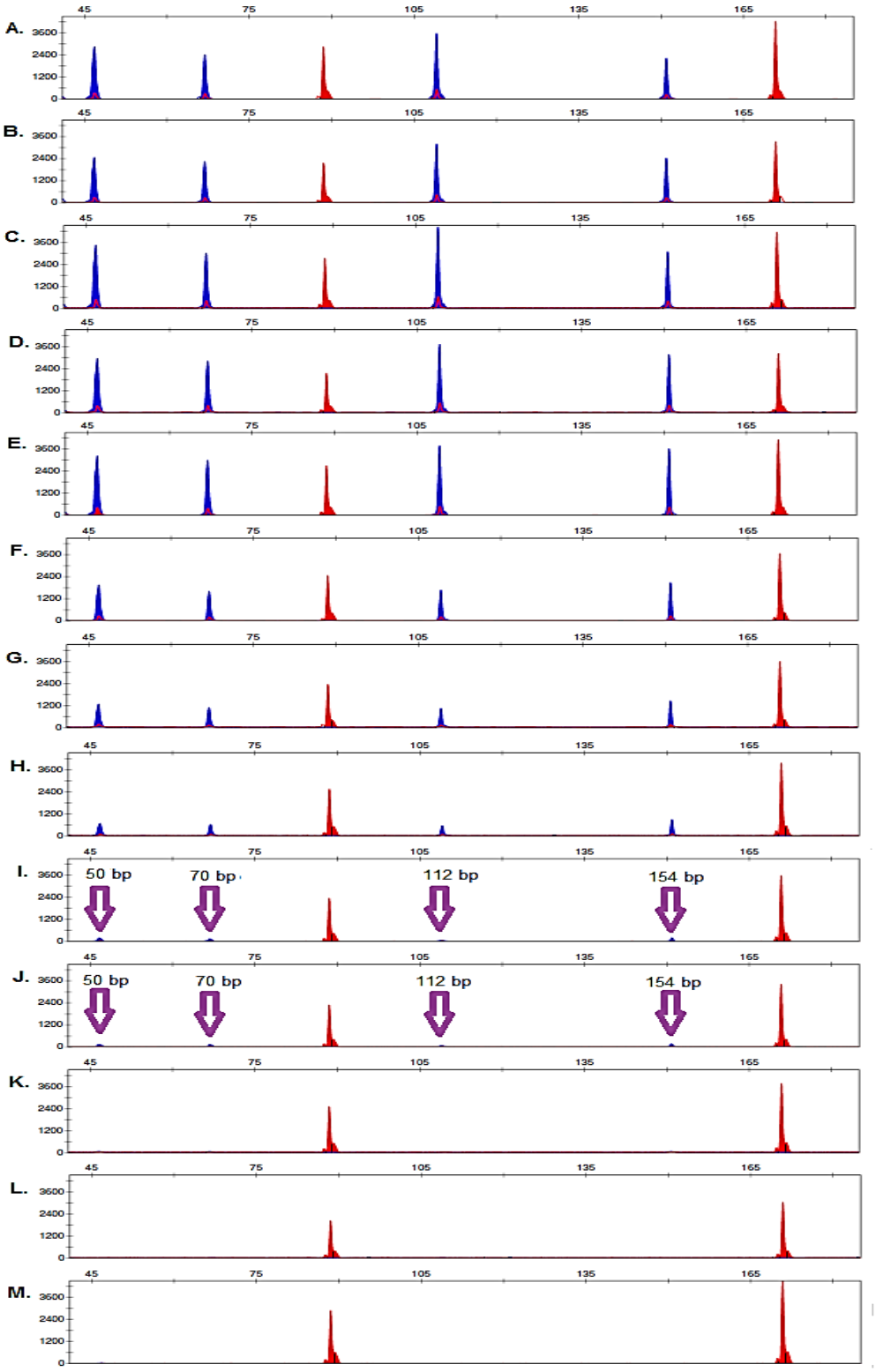


Figure 6-10: Electropherograms generated from serial diluted Human Genome control (Promega) with (A) 100 ng/µl, (B) 50 ng/µl, (C) 25 ng/µl, (D) 10 ng/µl, (E) 5 ng/µl, (F) 2.5 ng/µl, (G) 1.25 ng/µl, (H) 0.6 ng/µl, (I) 0.3 ng/µl, (J) 0.15 ng/µl, (K) 0.07 ng/µl, (L) 0.03 ng/µl (M) 0.01 ng/µl final concentrations. 1 µl of each sample was used for amplification. The blue peaks are the mini 4-plex which to detect degradation, while the red peaks are the IACs, to detect inhibition. The purple arrow shows dropout alleles.

6.3.2.3 Correlation graph plotting using the average peak heights and DNA concentrations

The whole data set was not suitable for a linear regression formation since the data has exponential distribution. Also, the data are not normally distributed thus no correction could be carried out to make a linear regression. The results of the statistical calculations are shown in Figure 6-11.

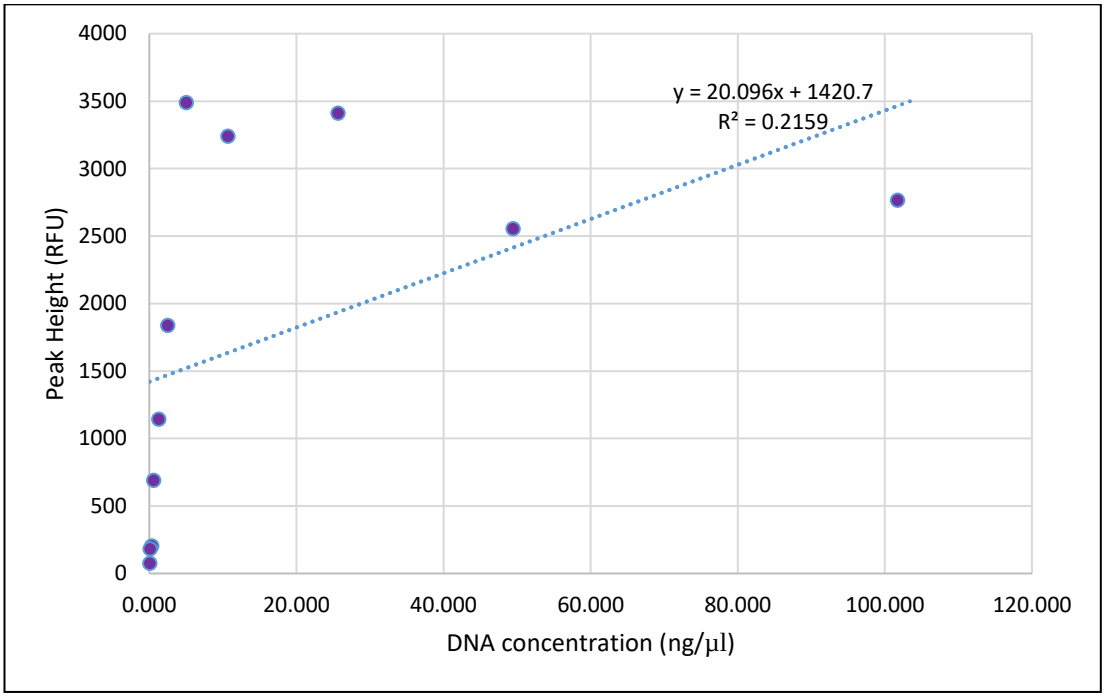


Figure 6-11 Statistical data of the comparison between average peak height of Mini 4-plex with different amount of DNA of 0.07 ng to 100 ng.

The range from 0.07 ng to 2.5 were selected for the correlation graph plotting. This is because degraded DNA usually fall into this range. The plotted graph is as shown in the Figure 6-12.

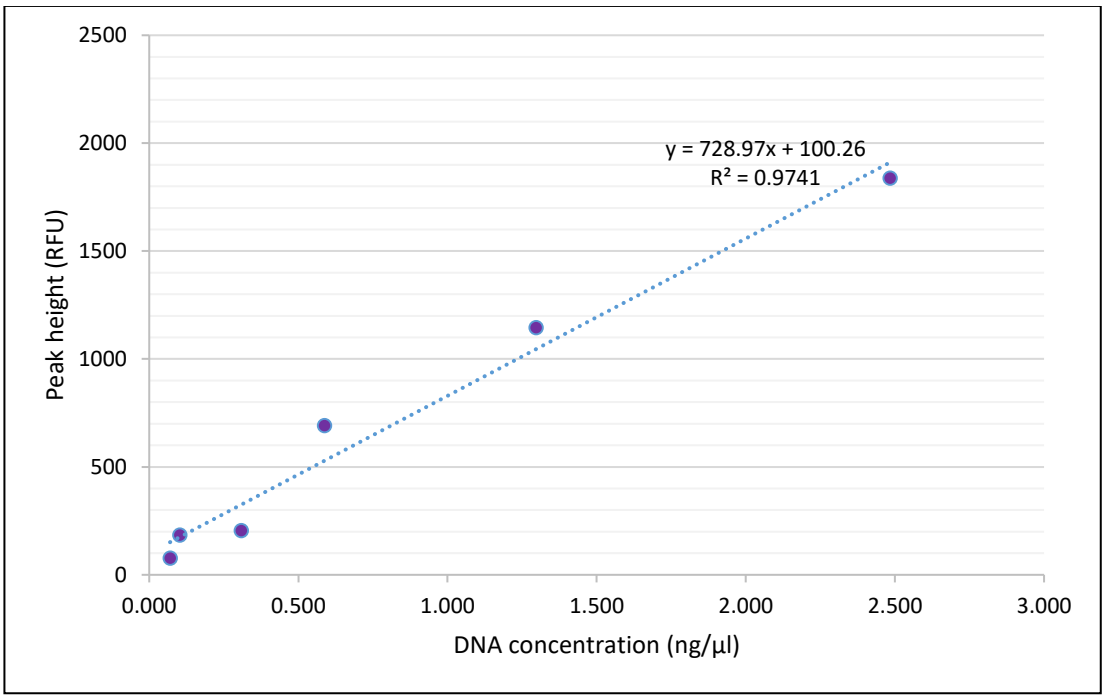


Figure 6-12: Statistical data of the comparison between average peak height of Mini 4-plex with different amount of DNA amounts of 0.07 ng to 2.5 ng.

The linear regression formula of y=728.97x+100.26 was obtained from the correlation graph. Also the coefficient of determination of (R^2) was 0.9741.

6.3.2.4 Evaluation of the correlation graph

Once the correlation graph was obtained, its efficiency to estimate the DNA concentration using the peak height was tested. This test was carried out using the same samples which were used to create this correlation graph (Table 6-7).

6.3.2.5 Concentration estimation using the Human Genome control.

For the estimation of DNA concentration using the correlation graph is useful for samples with DNA concentration between 0.07 ng/µl to 2.5 ng/µl were chosen since this correlation graph is useful for samples with DNA concentrations up to 0.07 ng/µl. The average peak heights were calculated using the linear regression formula (y=728.97x+100.26**)** to estimate the DNA concentration. The results are as shown in the Table 6-8.

Table 6-8 The comparison of the estimated DNA concentrations with their theoretical DNA concentration.

Average peak height (RFU)	Theoretical DNA concentration (ng/µl)	Estimation from correlation graph			
		DNA concentration (ng/µl)	Average of DNA concentration (ng/µl)	s.d	R.S.D (%)
2285	2.5	2.99	2.745	0.346	12.62
1630	1.25	2.10	1.675	0.601	35.88
1556	0.6	1.99	1.295	0.983	75.90
1126	0.3	1.41	0.855	0.785	91.80
1077	0.15	1.37	0.76	0.863	113.51
1085	0.07	1.35	0.71	0.905	127.48

Note: s.d,: standard deviation, R.S.D: Relative Standard Deviation, N/A: Not available

The estimated DNA results showed that this correlation graph is not very useful for DNA concentration estimation. The samples with concentration of 2.5 ng/µl, 1.25 ng/µl, 0.6 ng/µl, 0.3 ng/µl, 0.15 ng/µl and 0.07 ng/µl have R.S.D of 12.62%, 35.88%, 75.90%, 91.80%, 113.51% and 127.48% respectively showing that the estimated DNA concentrations are less precise. The inconsistent estimation among these DNA concentrations indicated that this correlation graph is not a useful tool for DNA concentration estimation.

6.3.2.6 Concentration estimation using casework samples

For further evaluation of the correlation graph, several bloodstains samples which were extracted using the Chelex® 100 were chosen as crime samples. These extracted samples were collected from KIMIA. Then, the samples were quantified using the Quantifiler® Human DNA Quantification kit (Applied Biosystems) and were amplified using the Mini-4plex PCR assay. The statistical data of this study are as shown in Table 6-9.

Table 6-9: The comparison of estimated DNA concentrations with their theoretical DNA concentration.

Average peak height (RFU)	Theoretical DNA concentration (ng/µl)	Estimation from correlation graph		
		DNA concentration (ng/µl)	s.d	R.S.D (%)
1027	1.16	1.27	0.08	6.40
2244	1.84	2.94	0.78	32.54
20417	1.57	27.87	18.60	126.32
2832	1.66	3.75	1.48	54.59
3183	1.16	4.23	2.17	80.55
5340	3.94	7.19	2.30	41.30
2791	1.86	3.69	1.29	46.63
2290	1.21	3.00	1.27	60.13
5340	3.94	7.18	2.29	41.21
784	0.57	0.94	0.26	34.65

Note: s.d,: Standard deviation, R.S.D: Relative Standard Deviation

The results showed that, only 1 out of 10 samples have R.S.D. below 10% thus good DNA concentration estimations for these samples were obtained. However, the success rate of this correlation graph for this batch of samples were only 10%. Even though unsatisfactory results were obtained, the efficiency of this correlation graph is based on the accuracy of each batch of analysis samples. If there is an error in the amplification process, differences in peak heights can occurred. So this estimation of the DNA concentration using this correlation graph becomes inaccurate. However, the evaluations of this correlation graph showed that it is not very useful for DNA concentration estimation.

6.3.2.7 Correlation graph of Internal Amplification Controls (IACs)

A graph was plotted between the average peak heights of IACs and the DNA concentrations to identify if there is any affect caused by the high amount of tested DNA on amplification efficiency of IACs. The average peak heights of IACs which were used to plot this correlation graph were obtained from the same electropherograms which were used to plot the correlation graph of 4-plex. The details obtained from the electropherograms are as shown in the Table 6-10.

Table 6-10 The average peak heights of IACs which were obtained from the electropherograms produced using the serial diluted DNA samples.

DNA concentration ng/µl	Average peak height (RFU)					
	Set A	Set B	Set C	Avg.	s.d.	R.S.D (%)
100	3554	3610	3521	3562	44.99	1.26
50	2744	2754	2787	2762	22.70	0.82
25	3424	3428	3744	3532	183.61	5.20
10	2683	2549	2716	2649	88.44	3.34
5	3410	3493	3540	3481	66.10	1.90
2.5	3060	3345	3469	3291	209.99	6.38
1.25	2463	2639	2722	2608	132.25	5.07
0.6	3255	3429	3454	3379	108.69	3.22
0.3	2966	2949	2663	2859	170.24	5.95
0.15	2850	2604	2477	2644	189.64	7.17
0.07	3120	3505	3150	3258	214.15	6.57
0.03	3122	3402	3313	3279	143.06	4.36
0.01	3732	3430	3548	3570	152.20	4.26

Note: Avg.: Average, s.d.: standard deviation, R.S.D.: Relative Standard Deviation.

The results showed that the similarity between each IACs of the same concentration is very high as the highest R.S.D. is only 7.17%. To confirm that IAC neither changed much when the tested DNA concentration increased, the correlation graphs using average peak heights of the IACs and the DNA concentrations was plotted. The plotted correlation graph is as shown in the Figure 6-13.

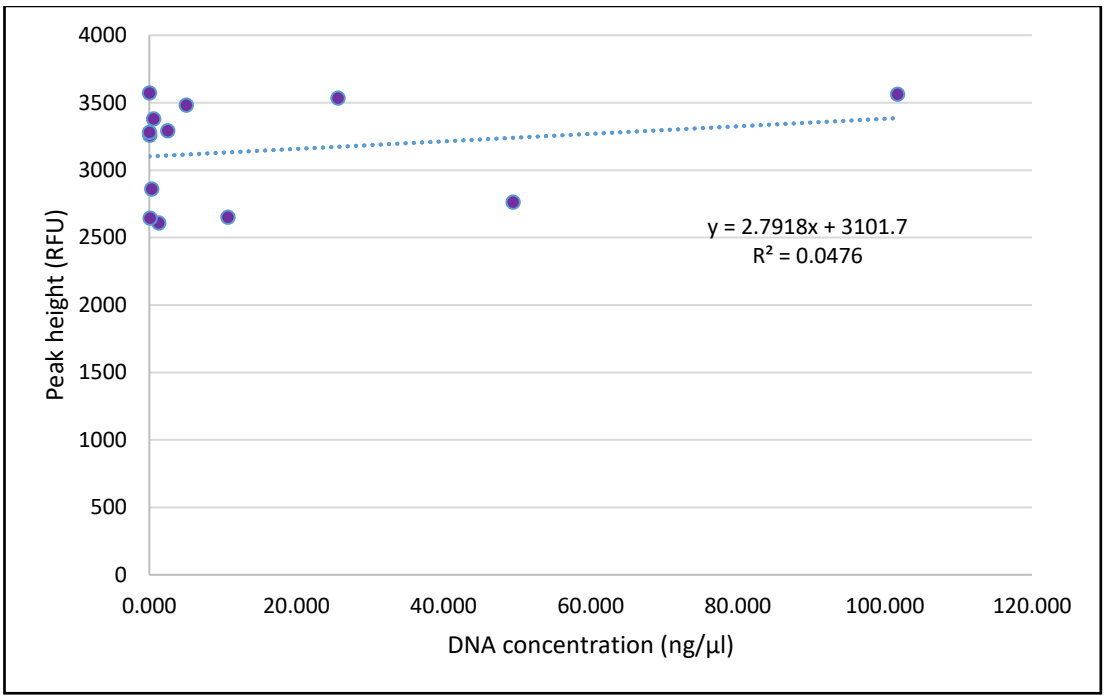


Figure 6-13 The correlation between average peak heights of IACs (IAC_{90} and IAC_{170}) with DNA concentrations.

The coefficient of determination (R_2) obtained from this correlation graph was as low as 0.046 showing that not much difference between the average peak heights of IACs which were developed in different DNA concentrations. This result showed that high amounts of tested DNA will not affect the amplification of the IACs. Thus, this multiplex system is not suitable for quantification purpose as the IACs only detect the PCR inhibitions and would not be affected by the DNA concentrations.

6.4 Discussion

6.4.1 Amplification of multiplex (Mini 4-plex & IACs) system on simulated forensic samples.

The aim of this study was to examine the effect of degradation and the detectable of the amplified primers using the multiplex (Mini 4-plex & IACs) system on simulated degraded samples in artificially or naturally degraded DNA at specified time intervals (e.g. hours, days, weeks, month and years).

Among the most challenging samples are those in which the DNA is highly degraded due to age or improper storage. The Quantifiler assay cannot be used as a precise measure of the extent of DNA degradation, but it can provide useful information on how much if any amplifiable DNA remains in a sample (Westring et al. 2007) That is why RSD values for all the saliva samples which exposed to various temperature at 3 hr and 48 hrs using distilled water and lysis buffer as wetting medium were high, between 19.66% to 63.97% and 10.91% to 62.96% respectively.

Both multiplex systems showed that it has a sensitivity level comparable to the leading commercial kits in the market (Gill, Curran & Elliot 2005). However, the Mini 4-plex is more sensitive and robust compared to the 4-plex. This is because through the analysis, the 4 plex showed immediate stochastic effect especially at larger amplicons. This is obvious at 48 hr using distilled water and lysis buffer. According to the model of random degradation of DNA, the amount of available template DNA decreases exponentially with increasing fragment size in damaged samples (Kitayama et al. 2013).

This approach allowed estimation of the number of detectable markers, and also useful for obtaining informative DNA profiles from a limited amount of DNA template. An alternative way to assess the degree of degradation and detect the presence of inhibitors in forensic samples is real-time PCR such as Quantifiler® Trio Kit as it can detect two human amplicons of 80 bp and 214 bp and degradation index could be used to study the quality and quantity of the samples.

6.4.2 DNA quantification using the multiplex (Mini 4-plex & IACs) system

Table below shows the advantages a limitation the usage of the Real Time PCR in molecular biology, much the same as in forensic DNA.

Table 6-11: The advantages and limitation of real time PCR (adapted from (Klein 2002)

Advantages	Limitations
Wide dynamic range of quantification (7 – log decades)	Product increase exponentially
High technical sensitivity (<5 copies)	Validation increases with cycle number
High precision (<2% of Ct values)	Increased variation after transformation to linear values
No post PCR steps	Overlap of emission[a]
Minimized risk of cross contamination	Maximal four simultaneous reactions[a]
High throughput	Increased risk of false negative results[b]
Multiplex approach possible	

[a]It is expected that new technology will improve this.

[b]In particular for pathogen detection (e.g. new emerging pathogens, highly variable pathogen)

Studies have shown that real time PCR would show DNA degradation by using the amplification data as the smaller fragments present a curve at lower cycles that the larger fragments. However, fragment analysis using Mini 4plex would be one way to assess how quickly and under what circumstances DNA breaks down, allowing analysts to look at the base pairs of the samples. The inhibitors can be detected by real-time PCR with internal amplification controls, but the quality of the samples cannot be predicted to obtain a good profile Plus, error in DNA quantification and normalization can affect the DNA profiles (Gill et al 2005).

The control samples which were quantified using the Quantifiler® Human DNA Quantification kit gave similar results and a more accurate result for the lowest DNA concentration tested (0.01 pg/µl) with R.S.D. below 5%. This is because the Quantifiler® Human DNA Quantification contains sequence specific DNA probes which are labelled with a reporter dye and emit fluorescence only once the probe has bound with its complementary DNA target. This further explains the accuracy and specificity of Quantifiler® Human DNA Quantification kit for human DNA quantification.

A correlation graph was created in this study utilizing the normal peak heights against the concentration of DNA template. Serial dilution of the Human Genome Control was used with the concentration of 0.01 ng/µl to 100 ng/µl. No artefact (e.g. split peaks) was obtained in all profiles even at higher concentration. The developed correlation graph has a linear regression formula of y=728.97x+100.26 and coefficient of determination

(R^2) of 0.9741. This correlation graph was developed using DNA concentrations in the range of 0.07 ng/µl – 2.5 ng/µl as this the range concentration of template DNA.

The success rate of only 10% was obtained using this correlation graph to estimate the concentration of an unknown DNA samples when the actual DNA concentration was between 0.07 – 2.5 ng and also the evaluations of this correlation graph showed that it is not very useful for DNA concentration estimation. Furthermore, this correlation graph cannot be applied on degraded DNA samples since the estimation is based on the average peak heights of the mini 4-plex and larger peaks will not be amplified with degraded samples. A previous study also showed that the peak height values sometimes show poor reproducibility even in standard DNA analysis (Manabe et al. 2013, Taylor et al. 2016), thus DNA concentrations may vary from the real quantity if estimated using a correlation graph.

In the meantime, the Internal Amplification Controls (IACs) were useful tools to detect inhibitors. The peak height has no interaction either with the amplification of Mini 4-plex or the IACs. This can be seen in the correlation graph of the average peak height for IACs against the DNA concentration produced the coefficient of determination (R^2) is 0.04. This means the variability of peak heights without any interference with the concentration of DNA.

6.5 Conclusion

The profiles generated using both in house multiplex; the 4-plex and Mini 4-plex on extracted saliva samples collected using distilled water and cell lysis (refer to section 6.3.1.2) showed significance difference especially at 48 h incubation in different temperature (20 °C, RT, 37 °C and 50 °C). Both multiplex systems showed that it has a sensitivity level comparable to the leading commercial kits in the market (Gill, Curran & Elliot 2005). However, the Mini 4-plex is more sensitive and robust compared to the 4-plex. This is because through the analysis, the 4 plex showed immediate stochastic effect especially at larger amplicons. This is obvious at 48 hr using distilled water and lysis buffer. In the other hand, the IACs are not affected by DNA quality as demonstrated in Sections 6.3.1.4 and 6.3.2.7.

Also, PCR inhibition can be detected through the use of an internal PCR control (IPC) during real-time quantification (Seo et al. 2012, Kontanis & Reed 2006). This technique could save more time and cost since less expensive reagents are involved while the Mini 4-plex & IACs multiplex has to be analyzed on a Genetic Analyzer to evaluate the quality of a sample which will caused more time and cost. Accurate quantification result also could be obtained from the real-time quantification, while the estimation of DNA quantification using this multiplex was very poor. As described in **Chapter 6**, a correlation graph was plotted based on the average peak heights of mini 4-plex and DNA concentrations from the serial diluted control DNA samples. This correlation graph has coefficient of determination (R2) of 0.9741 and can only be used when the sample concentrations are between 0.07 ng – 2.5 ng. Also, only 10% of tested samples produced relative standard deviations below 10% indicating inaccuracy of this correlation graph. However, there are limitations with real-time quantification approach where the amplicon in the PCR reactions is typically short, and so does not necessarily reveal the full extent of PCR inhibition. But with the novel mini 4-plex & IACs results, full extend of the sample quality could be revealed including the DNA degradation and the presence of PCR inhibition.

CHAPTER 7

DEVELOPMENT, OPTIMISATION AND VALIDATION OF A NEW MINI INDELS

7.1 Overview

It is a common problem in Forensic DNA laboratories to develop DNA profiles from highly degraded DNA cases; e.g. mass disaster where the DNA fragmented and modified resulted insufficient DNA could obtain for body identification. For example, Fondevila et al. reported a DNA identification case involving a set of skeletal remains found after a forest fire, which was later determined as belonging to a person missing for 10 years (Fondevila et al. 2008). However, in this case, they obtained insufficient DNA typing results even on using 3 short-amplicon short tandem repeat (STR) sets, and therefore had to identify the individual using single nucleotide polymorphism (SNP) analysis (Ozeki et al., 2015).

Research on multiplexes, have demonstrated that a reduction of target size significantly increases efficiency of amplification from highly degraded DNA samples (Butler et al., 2003, McCord et al., 2005, Opel et al., 2007, Shafique et al., 2016). STRs required a greater amount of template DNA with a higher molecular size (200 bp to 400 bp) for profiling compared to miniplexes (e.g. AmpFℓSTR® MiniFiler™ PCR Amplification Kit (Thermo Fisher Scientific) which used smaller sized PCR products less than 200 bp. Therefore, the possibility of obtaining full profiles with degraded DNA or samples in low amounts is higher.

The aim of this study, is to develop an amplification kit of Mini-INDELs by re-designing 9 amplicons from (Pereira et al. 2009) 38 biallelic markers in-house multiplex (B3, B4, B5, G6, G8, Y4, Y5, R3 and R5). The development of Mini-INDELs typing procedures include selection of marker, design of specific primer, amplification and optimisation of the mini-INDELs (Pereira 2012). The markers used in this research are chosen from NCBI dbSNP, based on the following criteria: (i) non-coding and diallelic, (ii) allele length variation of 2 to 6 bp (Oka et al., 2014, Murthy et al., 2015).

This part of research also aims to develop a Mini-INDEL multiplex by combining the 9 Mini-INDELs with another 15 Mini-INDELs (including amleogenin) which was developed

by UCLAN PhD student, thereby allowing for forensic applications. The loci include HLD67, HLD131, HLD6, HLD101, HLD124, HLD39, HLD58, HLD88, HLD99, HLD84, HLD97, HLD125, HLD128 and HLD 111) and sex identification locus amelogenin.

7.2 Objectives

- To develop a 9 Mini-INDEL assay and optimise the PCR system.
- To carry out validation of the 9 Mini-INDELs.
- To study the 9 Mini-INDELs multiplex performance by using 100 Malays blood stained specimen.
- To calculate the allele frequency and forensic indices (MP, PD and PE) of the 9 Mini-INDELs markers.
- To combine the 9 Mini-INDELs with 15 Mini-INDELs (Mini-INDEL multiplex) and develop an allelic ladder.

7.2 Materials and methods

7.2.1 Design of 9 Mini-INDEL primers

The primers for this study were chosen from (Pereira et al., 2009) previous study. Nine primers were identified; B3, B4, B5, G6, G8, Y4, Y5, R3 and R5. The basic criteria of the markers selection are the sizes of the amplicons; within 3 bp to 5 bp. These loci also showed heterozygosity values higher than 0.30 which confirmed that the selected INDEL loci are highly polymorphic and very informative among the three population groups (the African, European and East Asian) (Pereira et al., 2009). The positions of these markers were obtained using SNP ID# through NCBI website as shown in Table 7-1.

Table 7-1: Markers from published paper Pereira et al.,(2009) were selected for inclusion in the Mini-INDELs, including chromosomal position, location in the genome, rs number, motif, reference allele and expected range.

Locus	Chr	Position (bp)	rs number	Motif (+DIP)	Ref. allele	Expected range (bp)
B3	19	48896180	rs2307689	TTC	-DIP	76-79
B4	8	76681235	rs35769550	TGAC	-DIP	121-125
B5	22	25120901	rs2307700	TCAC	-DIP	84-88
G6	9	104626014	rs2307580	AATT	-DIP	106-110
G8	18	34677042	rs34511541	CTCTT	-DIP	72-77
Y4	16	83139788	rs2067208	GCCAG	-DIP	97-103
Y5	1	245878706	rs2307579	ATG	-DIP	75
R3	5	5178112	rs2307526	ACAC	-DIP	70-74
R5	10	54112392	rs1160886	ACT	-DIP	100-103

-DIP represents deletion

7.2.2 Primer design-based approach to multiplex design

These primers were designed complementary to the 15 mini-INDEL primers set which were reported by Majid Bashir (UCLAN PhD student). There are three basic factors e.g., allele range, fluorescent dye labels and amplification conditions, were considered in designing for the 9 Mini-INDEL assay. The primers were then run as singleplex on the Genetic Analyzer 3500 as described in Section 2.7.2.2. Amplification of the alleles were done with the primers. This gave an indication as to how the selected loci could be combined into the multiplex reaction. The 9 loci chosen were based on the reported gene diversity values, allele size ranges which were obtained from Rui Pereira's published paper. Primer BLAST software was used to re-design these primer sets. The self-complementarity of individual primers was assessed using the National Centre for Biotechnology Information (NCBI) Primer-BLAST (basic local alignment search tool) program (www.ncbi.nlm.nih.gov/Entrez/). Table 7-3 shows the 9 selected markers including the length, melting temperature (TM), percentage of G-C content (or Guanine-Cytosine content) and the primer pairs sequence from 5' to 3'.

Table 7-2: Dyed labelled primer sequences using in the 9 Mini-INDELs.

Marker		Length	TM (°C)	GC%	5' to 3'
B3	F	21	58.94	52.4	AAACCTCCTCCCTCAGAAGAG
	R	20	60.93	50.0	ATTGGGTATCCCCTCGTTTG
B4	F	21	58.76	47.6	ACTGCGTTTCTGTAGAGGAGT
	R	28	58.52	28.6	AACATCCAAATTAGCTTCACATTTTTTC
B5	F	19	59.81	57.9	TCAGTGAACCTGGGACAGC
	R	24	57.16	54.3	GGAATCCCCATTCTTTTACTGAC
G6	F	25	59.80	40.0	CACTAAACTCTTCTTCTTGCAGCTT
	R	22	57.58	45.5	GGAACAGGACCATAGCATAACT
G8	F	27	58.85	37.0	GGACTTTAGTAGAAGAGGAAAATACCA
	R	24	59.56	37.5	AAATGAGACCCTTCTTAGGTTCAA
Y4	F	19	58.38	57.0	GCCTGAATAGATGCACCCG
	R	24	58.06	37.5	AATGTTCTAGAATCCACAAAGAGC
Y5	F	27	57.42	29.6	ACCATTAATAATAAAGTGTGGAAAGAC
	R	31	56.61	25.8	TTTATAAATACATAAAACGTATGAGCTAAC
R3	F	25	56.48	36.0	CATAAAGCAACTCTATTCCTTTTCC
	R	28	60.64	39.2	AATAGTCCAGTCTACCCAAATGTATTCC
R5	F	25	57.64	36.0	CTGATTCTGTTCTTTGTAATTCGGA
	R	27	59.27	32.0	TTATATACAAAGCAAAGGATGCTCA

7.2.3 Assigning florescent labelled dyes

All primers were then organized for expected amplicon size and assigned to four different labelled dyes from Life Technologies™ (FAM = blue, NED = green, VIC = yellow and PET = red). The primers were distributed among four dyes to achieve a distance of at least 2 base pair between two loci (refer to Figure 7-1).

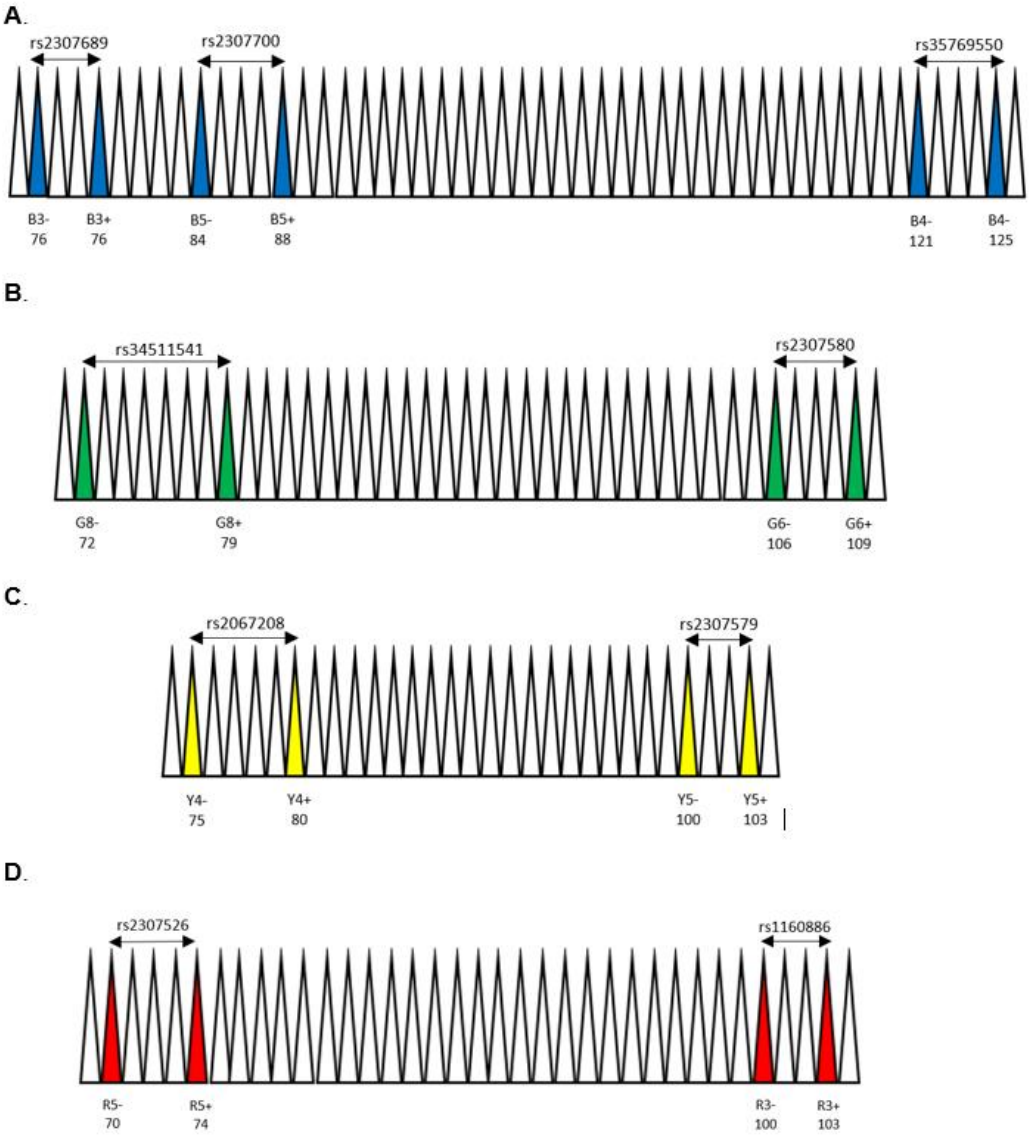


Figure 7-1: Size distribution of the mini INDELs in the four short amplicon sets: A) primers labelled with FAM; B) primers labelled with NED; C) labelled with VIC, and D) labelled with PET. The colours in the figure are arbitrary, and do not represent the actual fluorophore colours. The amplicons are presented by filled triangle, with the rs identification indicated on top and the alleles and base pairs indicated below.

7.2.4 Primer synthesis and purity

Primer pairs with 5' fluorescein labelled forward primers and unlabelled reverse primers were synthesized by Life Technologies, purified using HPLC and were delivered desalted and lyophilised. Stock solutions at 100 µM were prepared by adding the appropriate volume of the nuclease-free water and stored at – 20 °C. Aliquots of working solutions, 10 µM were prepared (10 µl from 100 µM to 90 µl of nuclease-free water) and stored at 4 °C.

7.2.4 Singleplex optimisation

Firstly, all primer pairs were initially optimised as singleplex in order to evaluate single primer efficiency and expected amplicon size. In order to optimise singleplex reaction, 13 µl of Platinum® PCR SuperMix High Fidelity (Thermo Fisher Scientific) was used for each reaction as described at Section 2.7.2.2. In addition, three different concentration of primers pairs, (0.3 µM, 0.5 µM and 1.0 µM) were used respectively with the master mix according to the optimised PCR condition (refer to Table 7-4).

Table 7-3: PCR conditions for amplification of the 9 Mini-INDEL individually using Platinum® PCR SuperMix High Fidelity (Thermo Fisher Scientific).

PCR stages		Temperature (ºC)	Time
Initial incubation		94	2 min
Denaturation	35	94	30 s
Annealing		54-60*	30 s
Extension		68	1 min
Hold		4	∞

* Annealing temperature = 54 °C, 56°C, 58°C, 60 °C.

7.2.5 Development of 9 Mini-INDEL assay

The new multiplex was developed using 9 pairs of primers. Thus, this multiplex amplifies B3, B5, B4, G8, G6, Y4, Y5, R5 and R5. The primer mix was prepared as shown in Table 7-5.

Table 7-4 Five primer sets for 9 Mini-INDELs PCR reaction tested for their optimum primer concentrations.

Primer	Forward and reverse primers	Annealing temp °C	Primer Concentration (μM)					Amplicon length
			Set A	Set B	Set C	Set D	Set E	
B3	AAACCTCCTCCCTCAGAAGAG	58.94	0.05	0.10	0.15	0.12	0.13	21
	ATTGGGTATCCCCTCGTTTG	60.93	0.05	0.10	0.15	0.12	0.13	20
B5	ACTGCGTTTCTGTAGAGGAGT	58.76	0.05	0.06	0.06	0.06	0.06	21
	AACATCCAAATTAGCTTCACATTTTTTC	58.52	0.05	0.06	0.06	0.06	0.06	28
B4	TCAGTGAACCTGGGACAGC	59.81	0.05	0.06	0.06	0.06	0.06	19
	GGAATCCCCATTCTTTTACTGAC	57.16	0.05	0.06	0.06	0.06	0.06	24
G8	CACTAAACTCTTCTTCTTGCAGCTT	59.80	0.05	0.06	0.06	0.06	0.06	25
	GGAACAGGACCATAGCATAACT	57.58	0.05	0.06	0.06	0.06	0.06	22
G6	GGACTTTAGTAGAAGAGGAAAATACCA	58.85	0.05	0.06	0.06	0.06	0.06	27
	AAATGAGACCCTTCTTAGGTTCAA	59.56	0.05	0.06	0.06	0.06	0.06	24
Y4	GCCTGAATAGATGCACCCG	58.38	0.05	0.06	0.06	0.06	0.06	19
	AATGTTCTAGAATCCACAAAGAGC	58.06	0.05	0.06	0.06	0.06	0.06	24
Y5	ACCATTAATAATAAAGTGTGGAAAGAC	57.42	0.05	0.10	0.15	0.12	0.13	27
	TTTATAAATACATAAAACGTATGAGCTAAC	56.61	0.05	0.10	0.15	0.12	0.13	31
R5	CATAAAGCAACTCTATTCCTTTTCC	56.48	0.05	0.10	0.15	0.12	0.13	25
	AATAGTCCAGTCTACCCAAATGTATTCC	60.64	0.05	0.10	0.15	0.12	0.13	28
R3	CTGATTCTGTTCTTTGTAATTCGGA	57.64	0.05	0.10	0.15	0.12	0.13	25
	TTATATACAAAGCAAAGGATGCTCA	59.27	0.05	0.10	0.15	0.12	0.13	27

7.2.5.1 Multiplex PCR reaction

After optimisation as singleplex reaction, the amplification of the 9 Mini-INDELs was performed in a single multiplex PCR reaction by using the following reaction as shown in Table 7-6.

Table 7-5: Several PCR reaction mix volumes used in this study to observe the sensitivity of this assay.

Component	PCR Reaction Mix Volume (µl/well)				
	MM1	MM2	MM3	MM4	MM5
Platinum® PCR SuperMix High Fidelity	12.5	12.5	5.0	12.5	10.0
Primer Mix	5.0	1.5	0.5	2.5	1.5
DNA template (0.5 ng/µl)	1.0	3.0	1.0	10.0	1.0
dH_2O	6.5	8.0	3.5	-	-
Final volume of master mix	25.0	25.0	10.0	25.0	12.5

In order to optimise the multiplex reaction, different parameters were used as below:

a) Primer concentration = 0.1 µM, 0.3 µM and 0.5 µM.
b) Input amount of DNA = 0.3 ng, 0.5 ng and 1.0 ng

Table 7-6: PCR conditions of 9 Mini-INDELs as a multiplex.

PCR stages		Temperature (°C)	Time (min)
Initial incubation		95	2
Denaturation	30	95	0.5
Annealing		58*	2
Extension		72	1
Final extension		60	30**
Hold		4	∞

* Annealing temperature = 54 °C, 56°C, 58°C, 60 °C

**Different final extension time from 30 to 80 mins were tested separately with the multiplex.

DNA amplification- the same PCR parameters were applied on the Mini-INDELs multiplex.

7.2.6 Development of allelic ladder

To facilitate consistent allele typing from one profile to another, allelic ladders were designed for each locus. The allelic ladders were designed using a combination of individual DNA samples which were representative of the most common alleles present in the populations (Butler 2011, Coble, Butler 2005). Amplification of the allelic ladder was performed using the same set of primers. This single amplification contained a range of reference amplicons which could be used for comparison when typing unknown samples. Two different concentrations; 0.5 ng and 1.0 ng of DNA mixed (of five individuals) were used to verify accurate allele designation and repeat structure.

Table 7-7 showing the amplicon sizes of both Mini-INDELs in development of allelic ladder.

InDel markers	Target sequence	Amplicon expected size (bp)
Blue INDEL		
B3-	3	76 - 79
HLD67+	8	74 - 92
B5-	4	84 - 88
HLD131-	11	86 - 97
B4-	4	91-95
X		106
Y		112
HLD6+	15	97 - 112
Green INDEL		
G8-	5	72 - 77
HLD101-	4	80 - 84
HLD124-	5	91 - 96
G6-	4	106 - 110
HLD39-	18	98 - 116
Yellow INDEL		
Y4-	5	75 – 80
HLD58+	4	83 - 97
HLD88-	9	90 - 99
Y5-	3	107 - 110
HLD99-	4	97 - 101
HLD84-	5	115 - 120
Red INDEL		
R5-	4	70 – 74
HLD97-	14	96 -100
HLD125-	6	89 - 95
HLD128-	8	99 -107
HLD111	4	80 - 84
R3-	3	110 - 113

7.2.8 Data analysis

The data obtained from the capillary electrophoresis (CE) were analysed using GeneMapper Software V2.1 (Applied Biosystems™). The panel and bin settings for the Mini-INDELs multiplex used in GeneMapper ID-X software V1.2 (Applied Biosystems™) are shown below in Table 7-8 and Table 7-9. These panels and bins were developed to analyse the data for ABI 3500 Genetic Analyzer. These panels and bins settings may vary to analyse data from other instruments like Genetic Analyzer 310 and 3130XL.

Table 7-8 Panel setting for Mini-INDEL makers including dye, colour, size (minimum and maximum) and ladder alleles.

Dye	Minimum size (bp)	Maximum size (bp)	Ladder Alleles
Blue	66	122	B3-, B3+, D67-, B5-, D131-, B5+, D67+, D131+, D6-, X, Y, D6+, B4-, B4+,
Green	62	133	G8-, G8+, D101-, D101+, D124-, D124+, G6-, G6+, D39-, D39+,
Yellow	68	118	Y4-, Y4+, D58-, D58+, D88-, Y5-, D88+, Y5+, D99-, D99+, D84-, D84+,
Red	65	115	R5-, R5+, D97-, D125-, D97+, D125+, D128-, R3-, R3+, D128+, D111-, D111+,

Table 7-9 indicating the bin settings for Mini-INDEL alleles for data generated from Genetic Analyzer 3500 including the bins' positions, dyes and widths.

INDEL Blue		
Marker	Bins position (bp)	Bin width (bp)
B3-	69.0	0.5
B3+	73.5	0.5
D67-	75.0	0.5
B5-	77.5	0.5
D131-	80.0	0.5
B5+	83.5	0.5
D67+	85.5	0.5
D131+	93.0	0.5
D6-	96.5	0.5
X	102.0	0.5
Y	108.0	0.5
D6+	111.5	0.5
B4-	115.0	0.5
B4+	119.5	0.5
INDEL Green		
G8-	64.5	0.5
G8+	72.0	0.5
D101-	75.5	0.5
D101+	81.0	0.5
D124-	84.0	0.5
D124+	91.5	0.5
G6-	101.5	0.5
G6+	106.0	0.5
D39-	110.5	0.5
D39+	131.5	0.5
INDEL Yellow		
Y4-	70.0	0.5
Y4+	76.0	0.5
D58-	77.0	0.5
D58+	82.5	0.5
D88-	84.5	0.5
Y5-	94.5	0.5
D88+	96.5	0.5
Y5+	97.5	0.5
D99-	101.0	0.5
D99+	108.5	0.5
D84-	110.0	0.5
D84+	116.5	0.5
INDEL Red		
R5-	67.0	0.5
R5+	74.5	0.5
D97-	76.5	0.5
D125-	86.0	0.5
D97+	92.0	0.5
D125+	95.0	0.5
D128-	96.0	0.5
R3-	100.5	0.5
R3+	103.5	0.5
D128+	105.5	0.5
D111-	106.5	0.5
D111+	112.5	0.5

7.3 Results

7.3.1 Optimisation of single PCR reaction

Amplification of the new markers were optimised using and Platinum® PCR SuperMix High Fidelity. PCR for each primer pair was set up individually using following PCR conditions: 94 °C for 2 min, followed by 94 °C for 30 secs, 58 °C for 30 secs, 68 °C for 1 min for 35 cycles and hold at 4 °C until samples were removed from the thermal cycler.

Promega Human Genomic DNA were used as a control in a PCR final volume 15 µl was used which containing: Platinum® PCR SuperMix (13 µl), 0.5 µl of appropriate concentration (0.5 µM) of each forward and reverse primer and DNA template 1 µl. Gradient PCR with annealing temperatures of 54 °C, 56 °C, 58 °C and 60 °C were used to test the primers using Veriti® 96-Well Thermal Cycler (Applied Biosystems) and ABI 2720 Thermal Cycler Life Technologies.

Three different primer concentration (0.3 µM, 0.5 µM and 1.0 µM) were tested to get optimum concentration. There was no obvious different of the peak morphology except for the peak height; yet all still above the threshold. Peak heights for G6- at 0.3 µM and 1.0 µM respectively, were a bit lower than G6+. Overall, the best peak height and quality were observed at 0.5 µM primer concentration.

The eight primer pairs were found to be optimum at all temperatures (54 ºC, 56 ºC, 58 ºC and 60 ºC) for B3, B4, B5, G6, G8, Y4, Y5, R3 and R5, respectively. The optimized primer pairs were assessed for any nonspecific amplification that would lead to extra peaks and could interfere with target peaks.

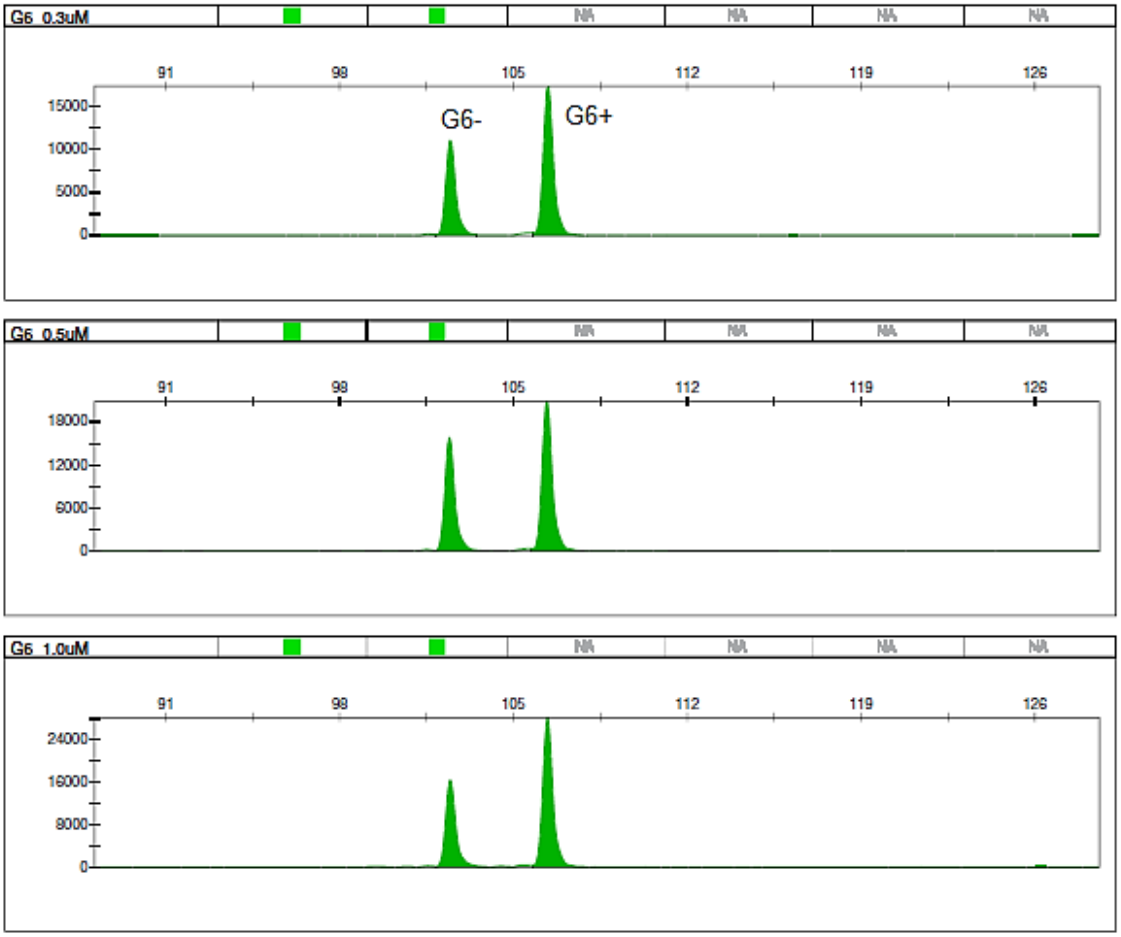


Figure 7-2: The electropherogram shows the optimisation results of the amplification of G6 marker in singleplex at three different concentrations (0.3 µM, 0.5 µM and 1.0 µM) and annealed temperature at 58 °C.

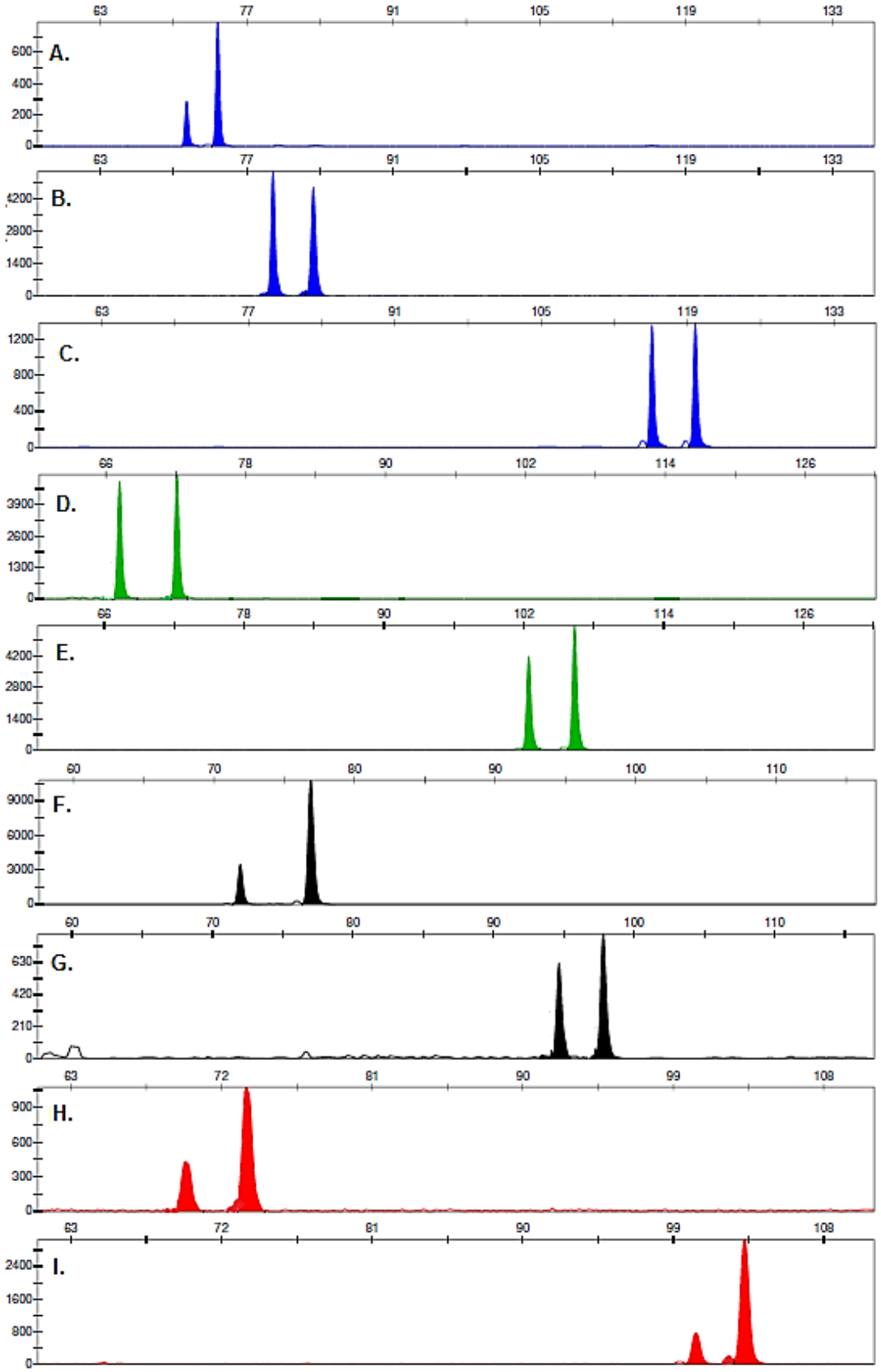


Figure 7-3: The nine panels of Mini-INDELs correspondence to FAM; (A) B3, (B) B5 and (C) B4; VIC; (D) G8 and (E) G6, NED; (F) Y4 and (G) Y5; lastly PET; (H) R5 and (I) R3 dye labelled peaks. The amplification of the 9 individual Mini-INDEL markers was set up using the annealing

temperature of 58 ºC and working concentration of 0.5 µM with Platinum® PCR SuperMix High Fidelity (Thermo Fisher Scientific).

7.3.3 Development of 9 Mini-INDELs multiplex.

After all the primers has been optimised, then the primers were combined in a multiplex. The primers were amplified using different sets of PCR volumes, primer concentration (0.05 to 0.4), primer mix concentrations (0.3 µM, 0.5 µM and 1.0 µM), DNA template (0.3 ng, 0.5 ng and 1.0 ng), annealing temperature (54 °C, 56 °C, 58 °C, 60 °C) and extension time (30 min to 70 min). The PCR products were genotyped on the ABI 3500 Genetic Analyzer.

Amplification of the Mini-INDELs as a multiplex was performed using Platinum® Multiplex PCR Master Mix (Thermo Fisher Scientific) and Human Genomic (Promega) was used as a control template in a PCR volume of 12.5 µl (Set 5) (Table)

The 9 primer pairs were found to works optimally at concentrations in Set E of 0.13 µM, 0.06 µM, 0.06 µM, 0.06 µM, 0.06 µM, 0.06 µM, 0.13 µM,0.13 µM and 0.13 µM for B3, B4, B5, G6, G8, Y4, Y5, R3 and R5, respectively (Table 7-10 and Figure 7-4). The primers then were mixed and 0.5 µM of primer mix is used.

Table 7-10 shows the concentration of the oligonucleotides as a primer mixture (total volume 18 µl) to amplify the mini-INDELplex PCR, prepared from 10 µM working solutions.

PCR Primers	Primer concentrations in a primer mix (µM)	Primer concentrations in a PCR final volume of 12.5 µl (µM)
B3	1.1	0.13
B4	0.5	0.06
B5	0.5	0.06
G8	0.5	0.06
G6	0.5	0.06
Y4	0.5	0.06
Y5	1.1	0.13
R5	1.1	0.13
R3	1.1	0.13

In order to get the optimised quantity of DNA, three different amounts were used; 0.3ng, 0.5 ng and 1.0 ng. Peak height at 0.3 ng was decreased, especially at red dyed loci, compared to 0.5 ng and 1.0 ng. However, there was no artefacts obtained at 1.0 ng even though the input is larger, but the RFU was high. The optimum DNA template to be used in this assay is 0.5 ng.

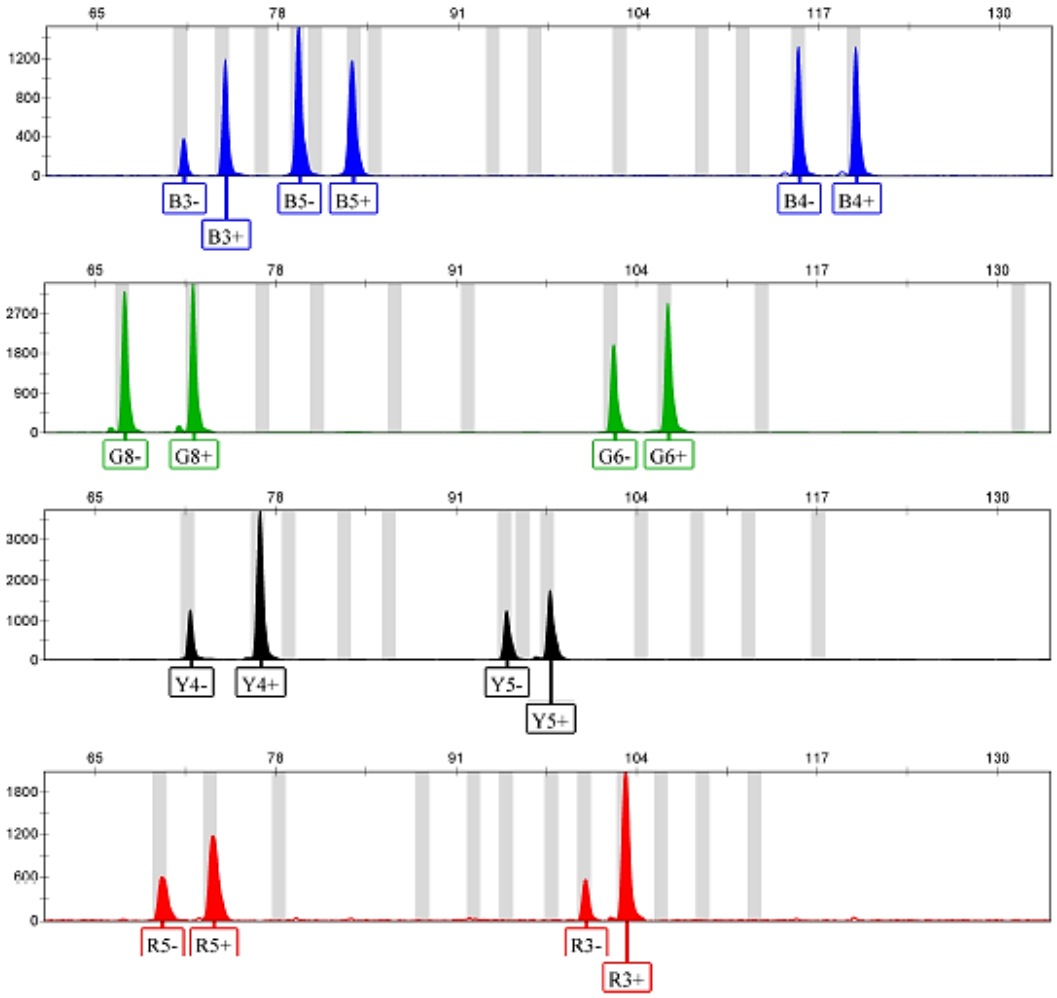


Figure 7-4: Representative electropherograms showing the profile of 0.5 ng of control Human Genome Control (Promega) amplified with the 9 Mini-INDELs. The genotype is shown with the allelic displayed underneath each peak.

7.3.3.1 Annealing temperature and extension time.

Thermal cycler conditions were optimised in terms of annealing temperature; 54 °C, 56 °C, 58 °C, 60 °C. Good results were obtained for all temperature (see Figure 7-5). However, 58 °C were chosen because to comply with the primers of 15 Mini-INDELs multiplex which reached optimum annealing temperature at 58 °C. Extension time for 30 min gave optimum results and no split peaks were observed due to incomplete adenylating (see figure).

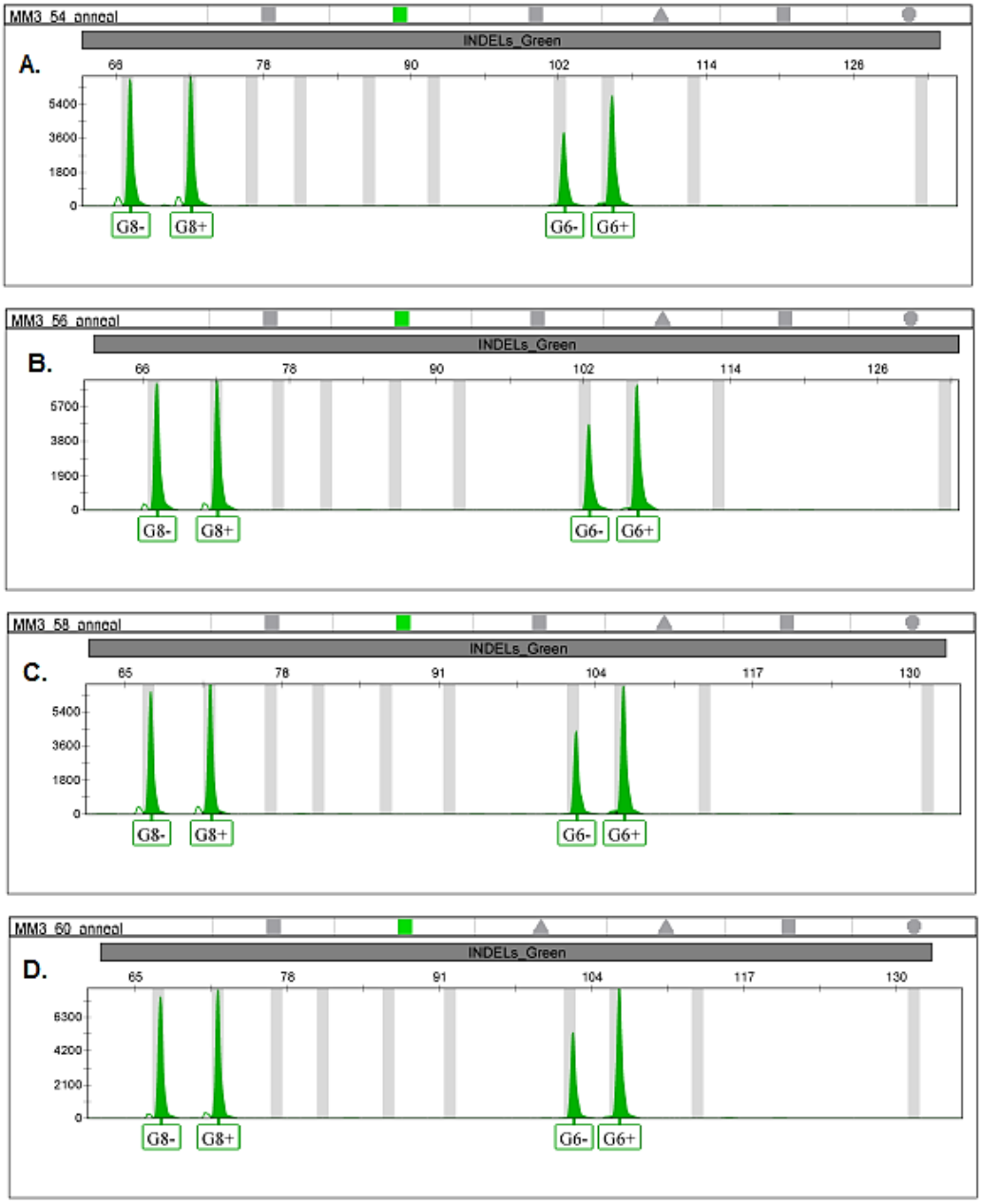


Figure 7-5 The electropherogram of the mini- INDEL multiplex amplification results for the markers G8 and G6 at different annealing temperature A) 54 °C, B) 56 °C, C) 58 °C and D) 60 °C with primer concentration 0.5 µM and 0.5 ng DNA of DNA template.

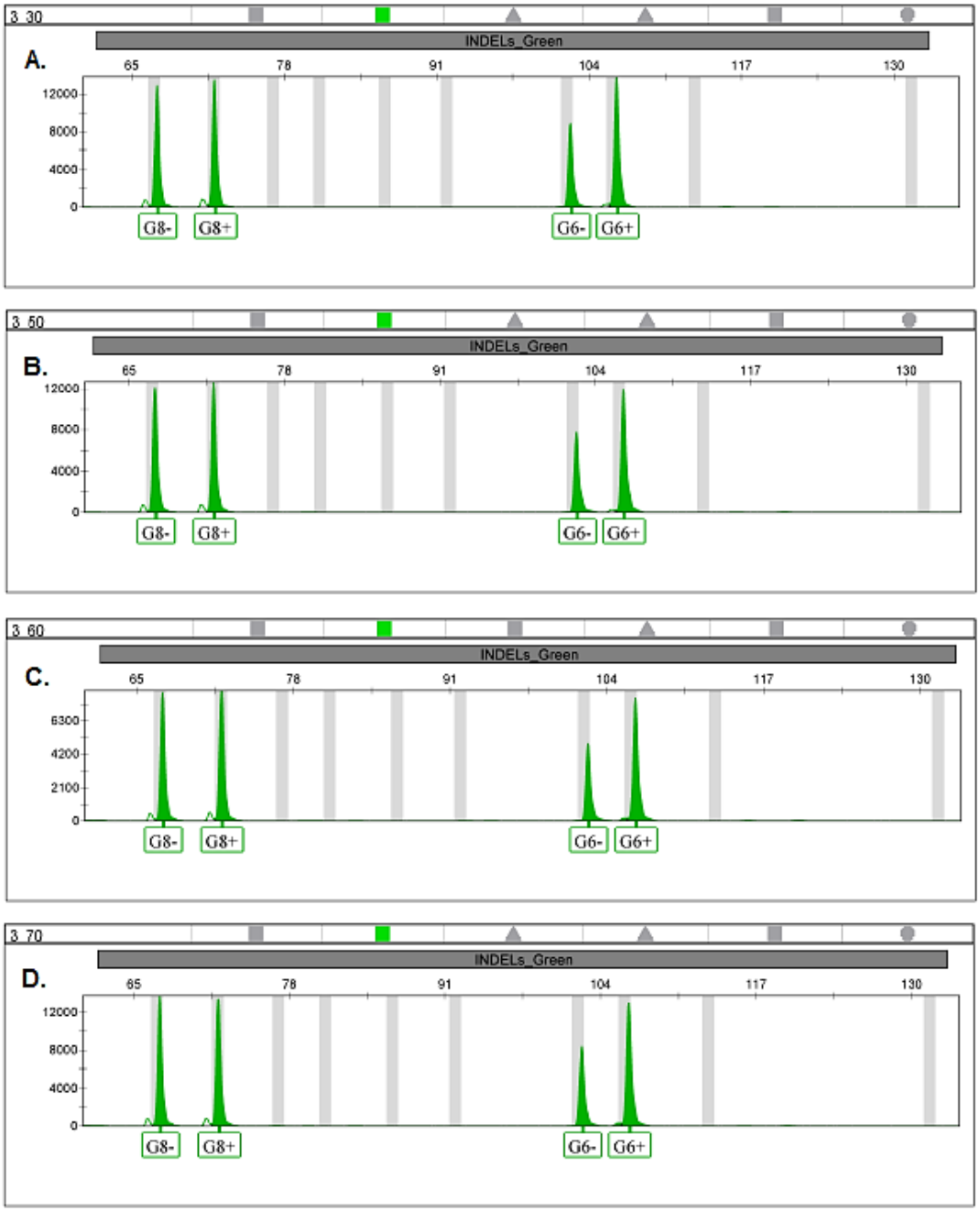


Figure 7-6 shows the electropherogram of the mini- INDEL multiplex amplification results for the markers G8 and G6 at different extension time A) 30 min, B) 50 min, C) 60 min and D) 70 min with primer concentration 0.5 µM and 0.5 ng of DNA template.

7.3.3.4 Reproducibility

To test genotyping consistency, random population samples were selected to amplified using the Mini-INDELs in two different thermal cyclers; Applied Biosystems Veriti™ 96-Well Thermal Cycler and Applied Biosystems 2720 Thermal Cycler (Life Technologies. The samples were carried out in duplicate as mentioned in Chapter 2. Consistent genotypes were obtained from all samples tested. The DNA profiles generated were identical showing the system is robust and results were reproducible. Figures 7-7 until 7-11 shows electropherogram for the Malay, Chinese, Indian, Iban and Bidayuh respectively.

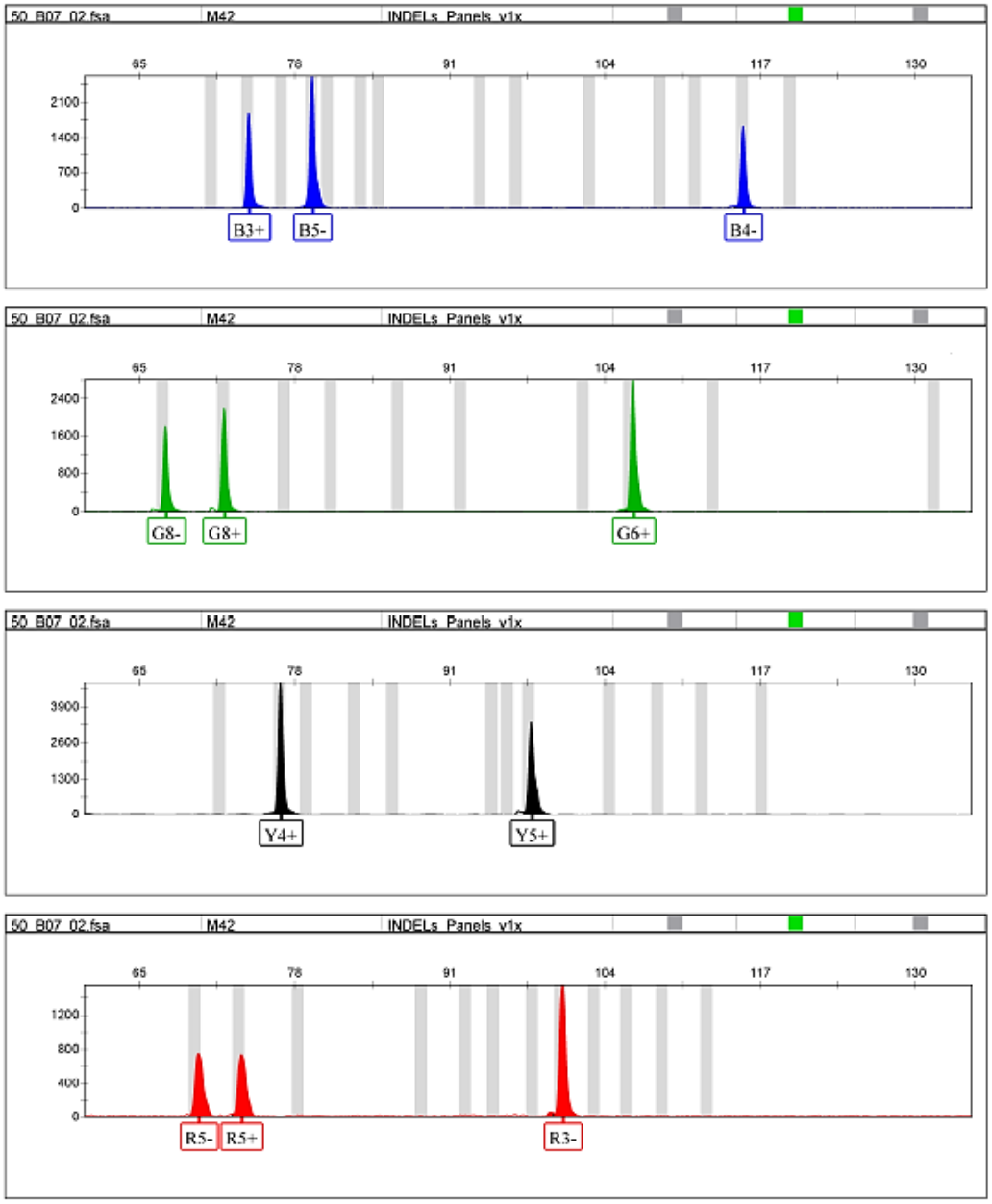


Figure 7-7: A representative profile of the Malay population profile using 9 Mini-INDEL assay.

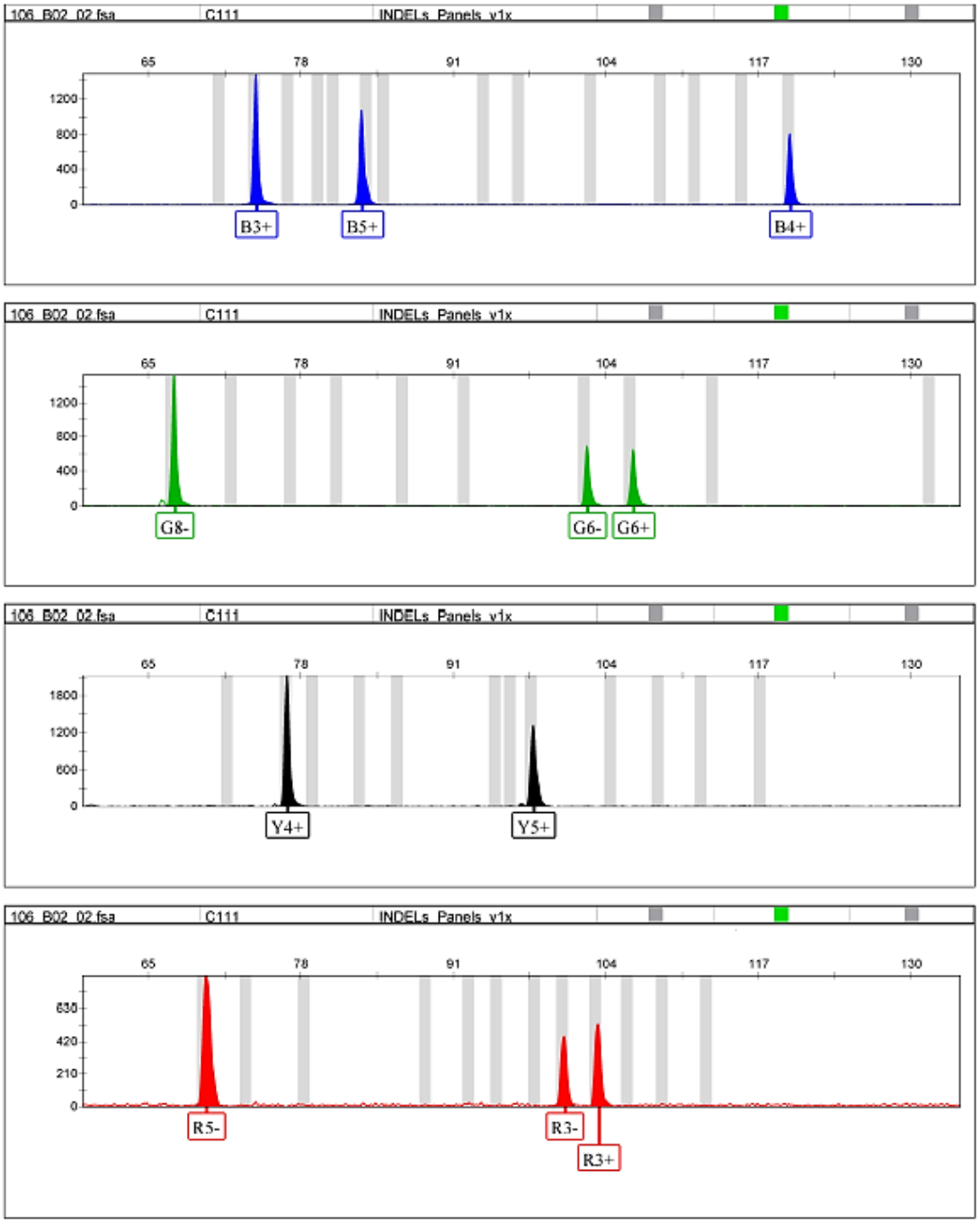


Figure 7-8: A representative profile of the Chinese population profile using 9 Mini-INDEL assay.

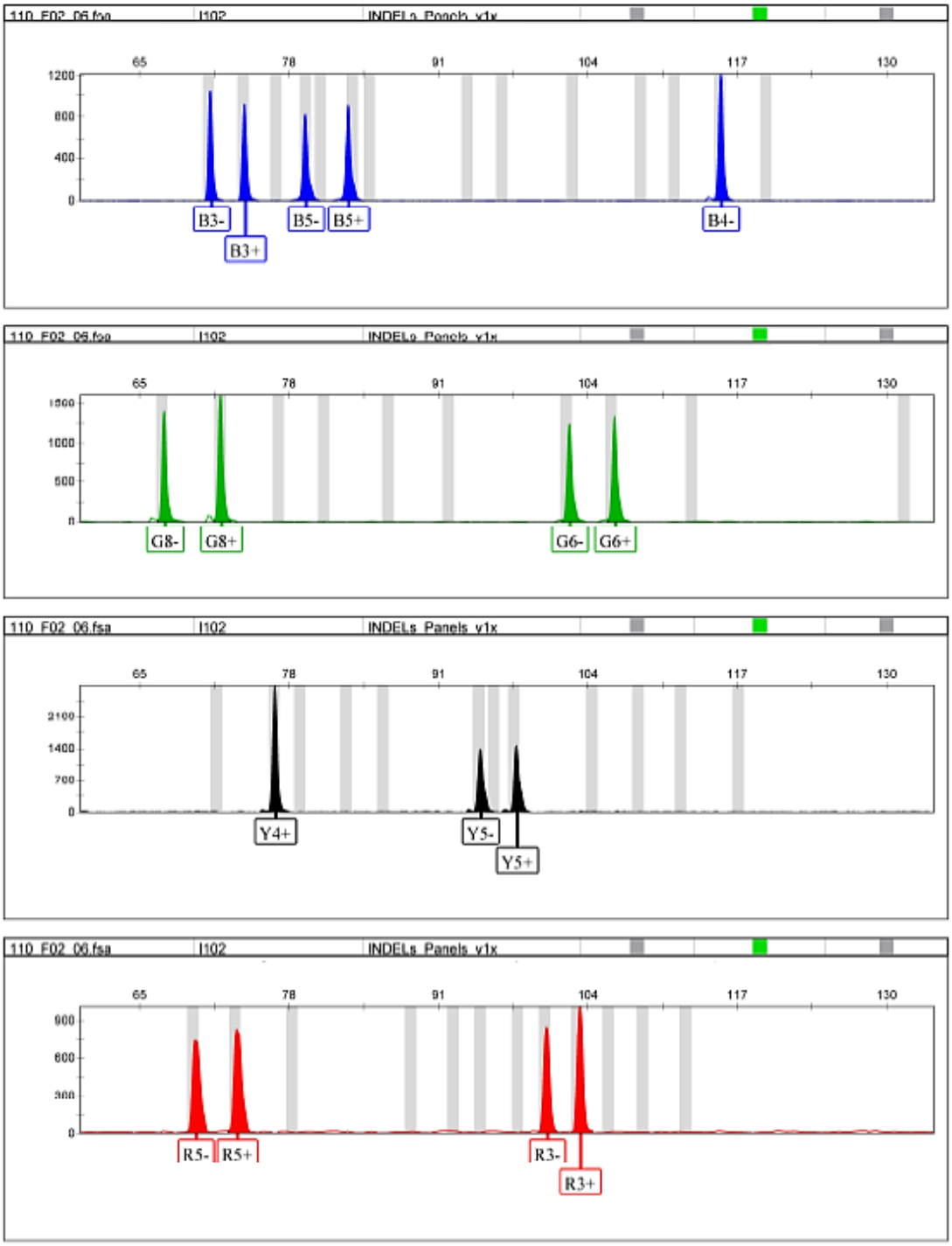


Figure 7-9: A representative profile of the Indian population profile using 9 Mini-INDEL assay.

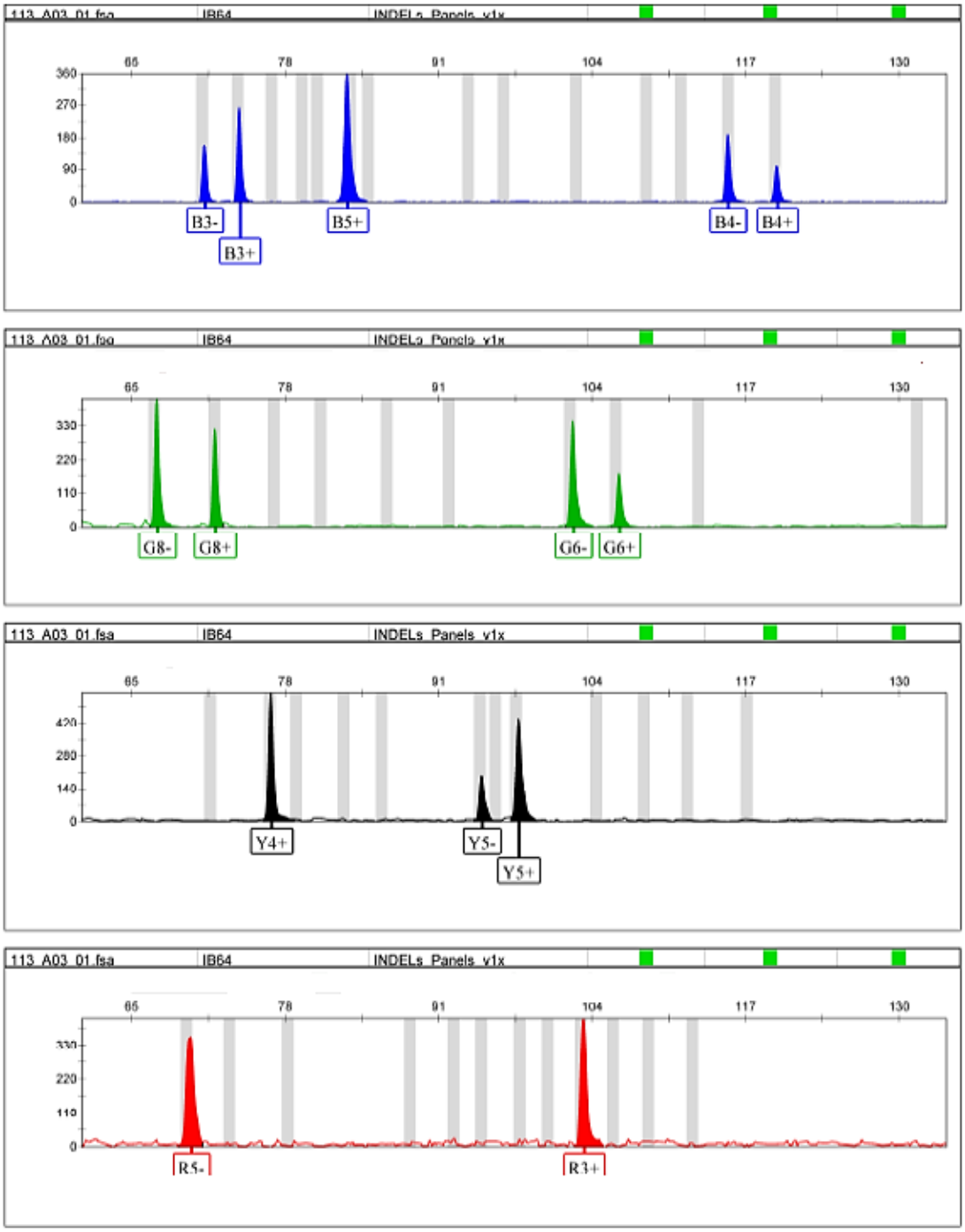


Figure 7-10: A representative profile of the Iban population profile using 9 Mini-INDEL assay.

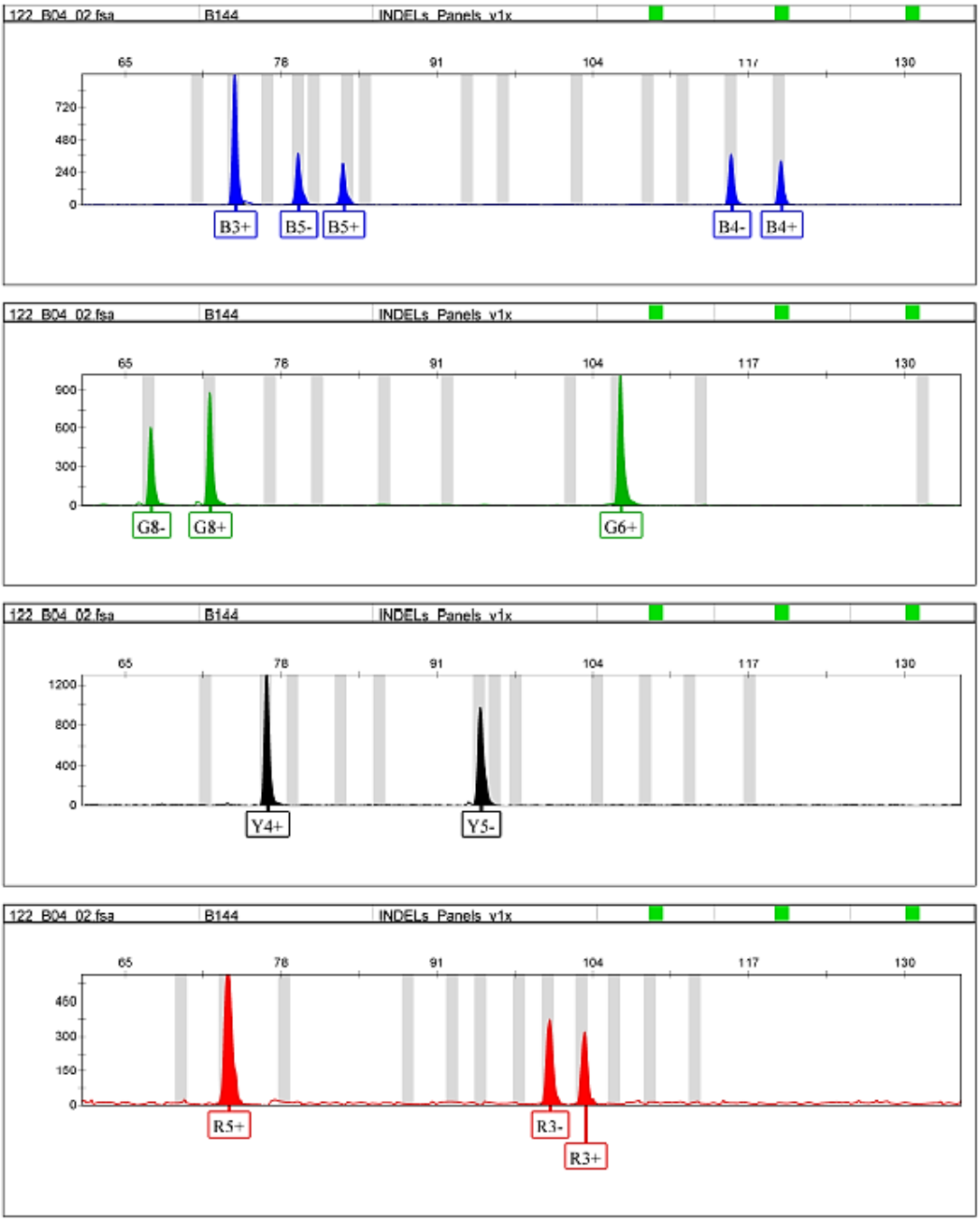


Figure 7-11: A representative profile of the Bidayuh population profile using 9 Mini-INDEL assay.

7.3.5 Statistical value based on 9 Mini-INDELs

Out of the 105 Malay subpopulations samples, 5 could not be amplified and were therefore not included for further analysis. The database was then compiled of the DNA profiles of 100 reference specimen. The samples were typed with the Mini-INDELs I multiplex and 9 INDEL loci were analysed.

7.3.5.1 Allele Frequency

Table 1 showed the allele frequencies obtained for the Malays subpopulation group. Allele frequencies for the short (deletion; which was coded as **1**) and long (insertion; which was coded as **2**) alleles have been calculated for the Malay, comprising of 100 individuals. Allele frequencies of the 9 Mini-INDEL markers are listed in Table 7-11. Thus, the minimum allele frequency observed is 0.140 (–Y4). Whilst, the largest allele frequency observed in Malays is 0.860 (+Y4).

Table 7-11: Allele frequency for Malay population

Mini-INDEL Locus	Allele frequencies (-/+)
B3	0.290/0.710
B5	0.425/0.575
B4	0.585/0.415
G8	0.375/0.625
G6	0.620/0.380
Y4	0.140/0.860
Y5	0.260/0.740
R5	0.530/0.470
R3	0.455/0.545

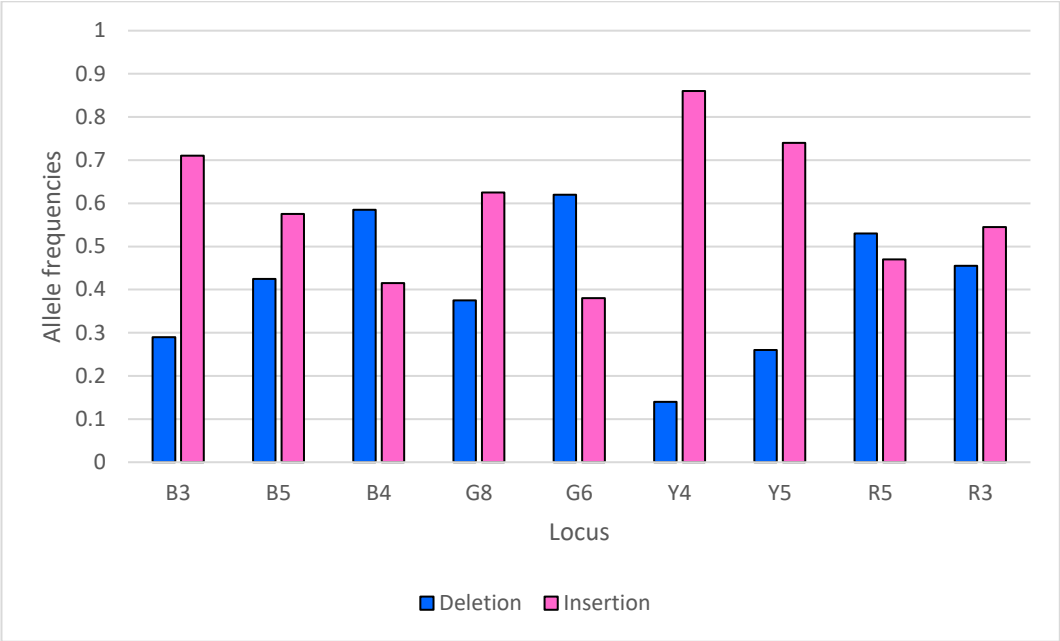


Figure 7-12 Graph of allele frequencies of the Malays subpopulation for 9 Mini-INDEL loci. The green blue bar represents deletion, while the pink represents insertion.

7.3.5.2 Forensic parameters

The forensic suitability of the 9 Mini-INDEL loci was evaluated for the Malays subpopulation. The combined power of discrimination (CPD) for the 9 Mini-INDEL loci was 0.992 for the Malays and the Combined Match Probability (CMP) was 3.06×10^{-4}. This implies that when using these loci combined in the studied population a high degree of discrimination could be determined.

Single locus power of exclusion (PE) values across all nine loci ranged between 0.035 to 0.188 (Table). This indicates a low degree of exclusionary power when using the loci individually. However, when combining the PE values in the Malay population, the forensic ability of these loci are markedly increased. The Combined Probability of Exclusion (CPE) calculated was 0.976435.

The PIC value is 1; meaning that the polymorphism is informative in any random mating. As shown in Table, the PIC of 8 out of 9 Mini-INDEL loci proved to be greater than 0.25, which indicates that B3, B5, B4, G8, G6, Y5, R5 and R3 were highly polymorphic in the investigated Malay subpopulation. The most polymorphic loci were B5, B4, R5 and R3 (0.37 respectively) and the least polymorphic one was Y4 (0.21).

Table 7-12: Forensic efficiency parameters of the 9 loci in 100 unrelated Malay individuals.

Mini-INDEL Locus	Match probability (MP)	Power of discrimination (PD)	Power of exclusion (PE)	PIC
B3	0.425	0.575	0.102	0.33
B5	0.373	0.627	0.163	0.37
B4	0.362	0.638	0.133	0.37
G8	0.367	0.633	0.097	0.36
G6	0.379	0.621	0.140	0.36
Y4	0.612	0.388	0.035	0.21
Y5	0.455	0.545	0.114	0.31
R5	0.377	0.623	0.188	0.37
R3	0.365	0.635	0.163	0.37

7.3.6 Combination of 9 Mini-INDEL and 15 Mini-INDEL multiplex

A few of the initially selected primer pairs were not suitable for incorporation into the multiplex due to overlap with neighbouring markers or the distance is too near (approximately 1 bp).

7.3.6.1 Redesigning of B4, Y5 and R3 reverse primers

During the analysis of some profiles generated from New Mini-INDELs assay, it was observed that four pairs of alleles were overlapping with each other; D67+ with B4- and D111+ with R3+. While alleles Y5- and D99- were one base pair apart which resulted a split peak (see figure). To overcome these problems, reverse primers for B4, Y5 and R3 were redesigned and optimised again. After new primers were introduced to the multiplex, the overlapped alleles are separated and resulted into distinctive peaks (see figure).

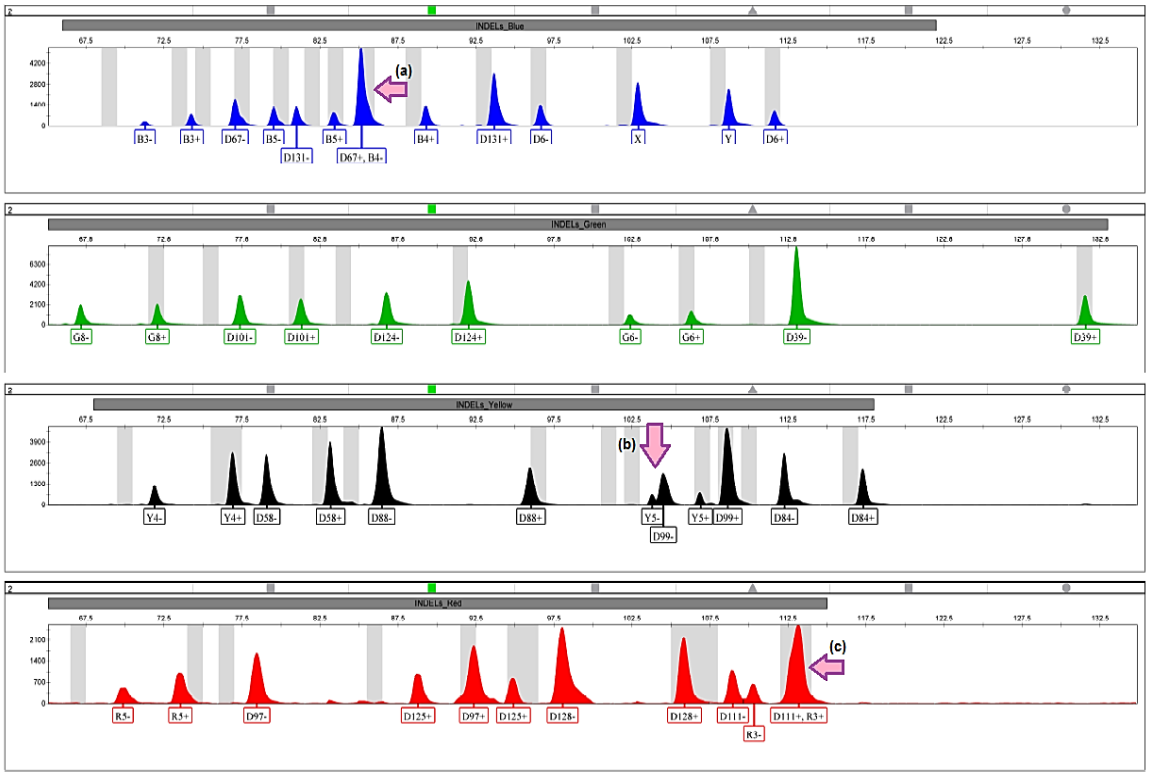


Figure 7-13 Alleles overlapping at (a) D67+ with B4- and (c) D111+ with R3+, and (b) Y5- is one base apart with D99- which developed a split peak.

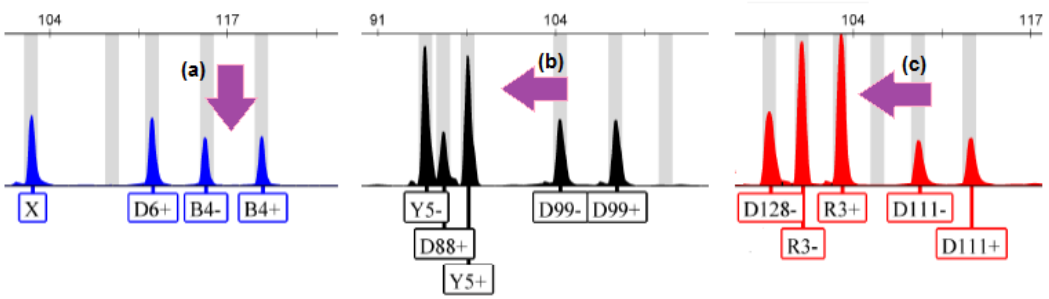


Figure 7-14: After redesigned and rearrangement of the reverse primers of B4, Y5 and R3 respectively, the peaks are assigned correctly. Note: the primers which mentioned before are the new designed primers

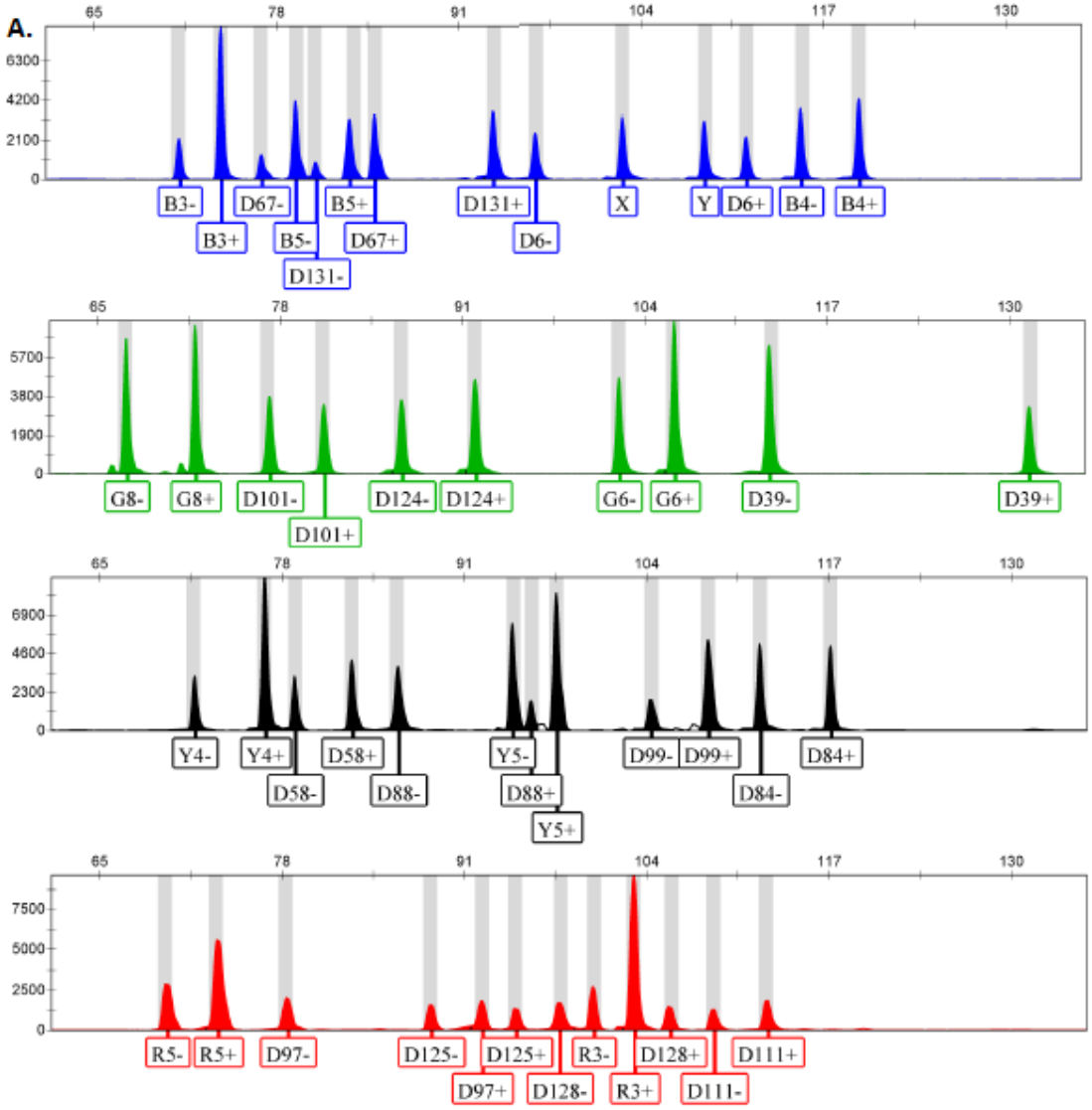


Figure 7-15 (A): The Mini INDEL systems produced a profile using extracted Human Genomic Male DNA(Promega) which is as control DNA in this reaseach.

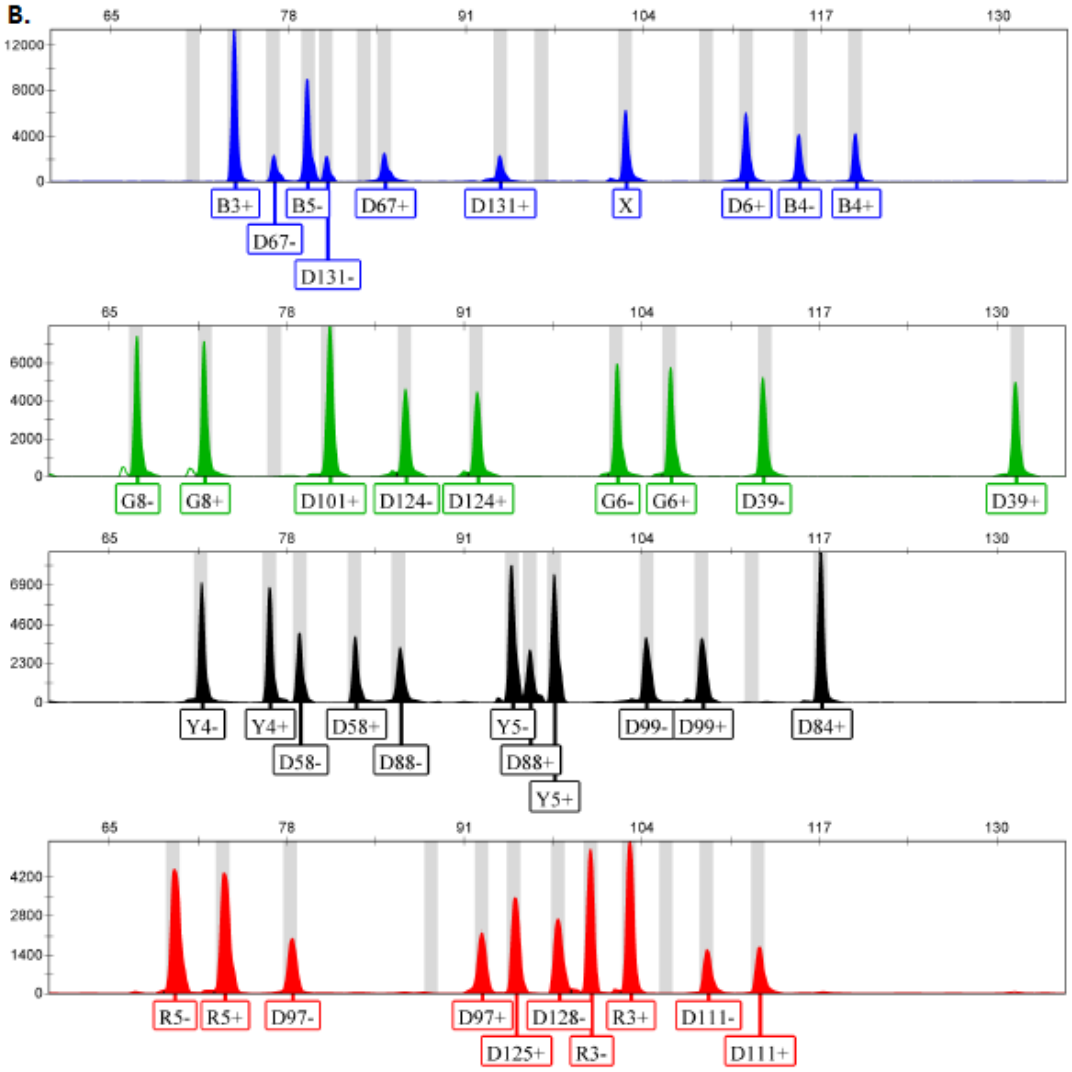


Figure 7-16: The Mini INDEL system produced profile using extracted buccal swabs sample which used as reference sample in this reasearch. This profile shows a female (XX) profile.

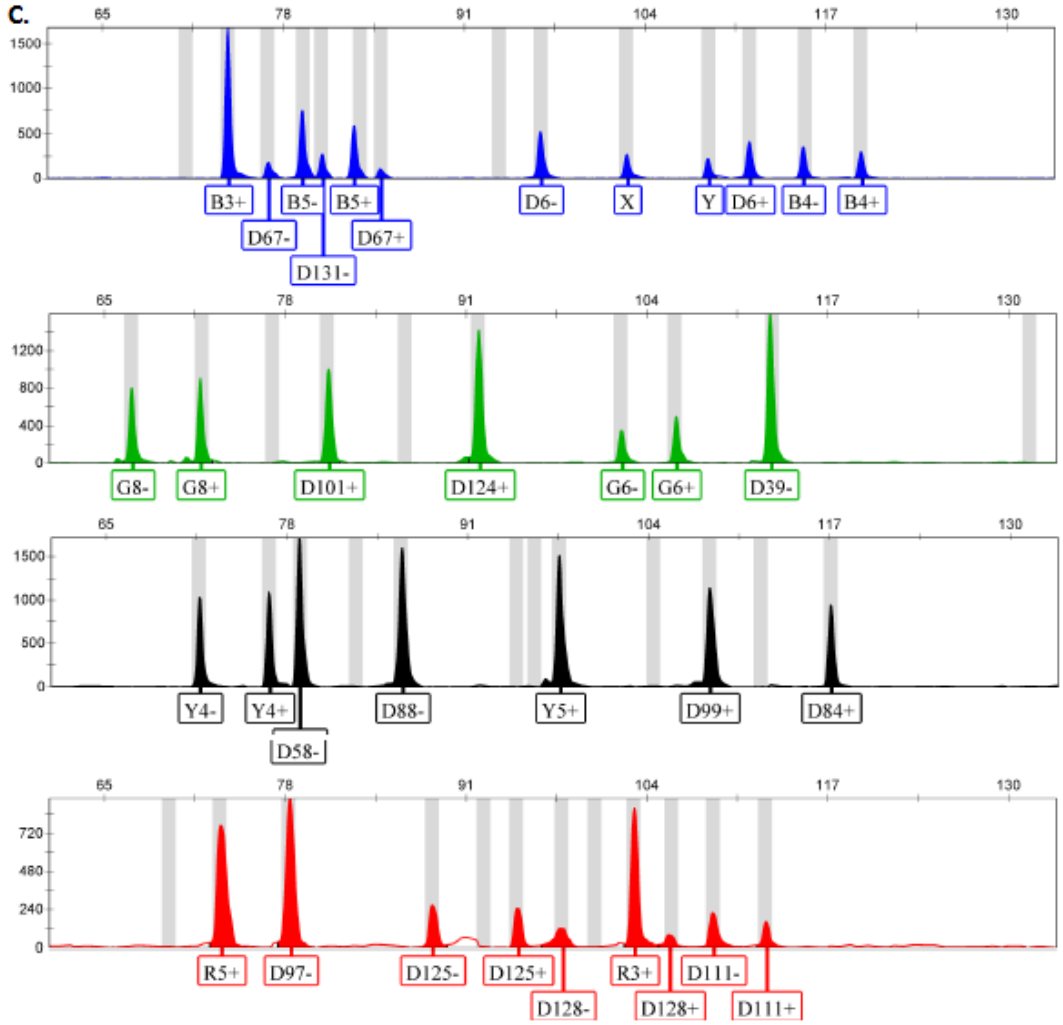


Figure 7-17: The Mini INDEL system produced profile using extracted buccal swabs sample which used as reference sample in this reasearch. This profile shows a male (XY) profile.

7.3.6 Development of allelic ladder

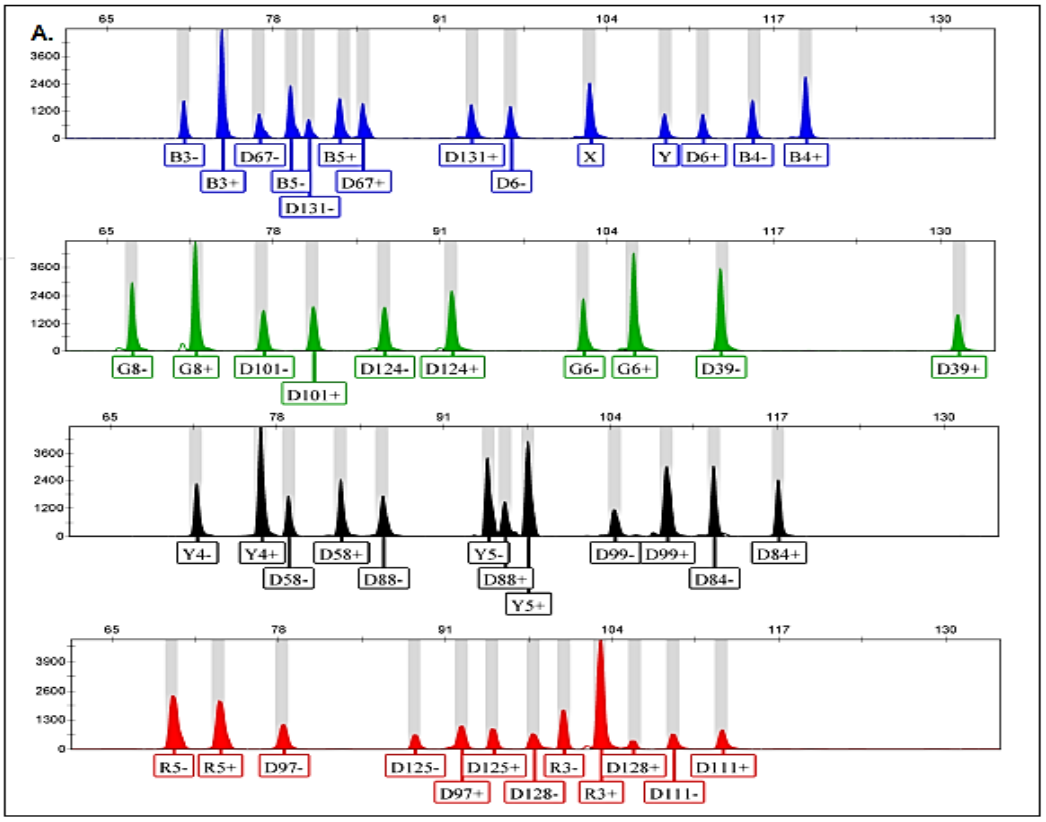
Subsequent to the design of allelic ladders, mini INDEL analysis of individual samples were conducted to ensure that correct allelic designations were made. Allelic ladder for the loci B3, HLD67, B5, HLD131, B4, Amelogenin, HLD6, G8, HLD101, HLD124, G6, HLD39, Y4, HLD58, HLD88, Y5, HLD99, HLD84, R5, HLD97, HLD125, HLD128, HLD111 and R3 were successfully generated. All polymorphic forms of alleles for the specific locus were mixed and balanced together making a final allelic ladder. Each alleles creating a gap that simplifies identification of the alleles. Two different concentrations of DNA mixed (0.5 ng and 1.0 ng) were tested (see Figure). This allelic ladder allows confident, rapid and precise assignment of discretely defined alleles. Table 7-12 represents allelic ladder expected and observed amplicon sizes for each allele. However, the alleles were bit shifted from the expected allele.

Table 7-13 The difference of expected observed allele in the multiplex.

Indel Markers	Amplicon expected size (bp)	Amplicon size on observed EPG in Mini-INDEL Multiplex (bp)
Blue INDEL		
B3-	76	71
B3+	79	74
D67-	74	77
B5-	84	79
D131-	86	81
B5+	88	83
D67+	92	85
B4-	121	115
B4+	125	119
D131+	97	93
D6-	97	96
X	106	102
Y	223	108
D6+	112	111
Green INDEL		
G8-	72	67
G8+	77	72
D101-	80	77
D101+	84	81
D124-	91	86
D124+	96	91
G6-	106	102
G6+	110	106
D39-	98	113
D39+	116	131

Table 7-14 The difference of expected observed allele in the multiplex (continue).

Yellow INDEL		
Y4-	75	71
Y4+	80	76
D58-	83	79
D58+	97	83
D88-	90	86
D88+	99	95
Y5-	100	94
D99-	97	104
Y5+	103	97
D99+	101	108
D84-	115	112
D84+	120	117
Red INDEL		
R5-	70	69
R5+	74	73
D97-	96	78
D125-	89	88
D97+	100	92
D125+	95	94
D128-	99	97
D128+	107	105
D111-	80	108
R3-	100	100
D111+	84	113
R3+	103	103



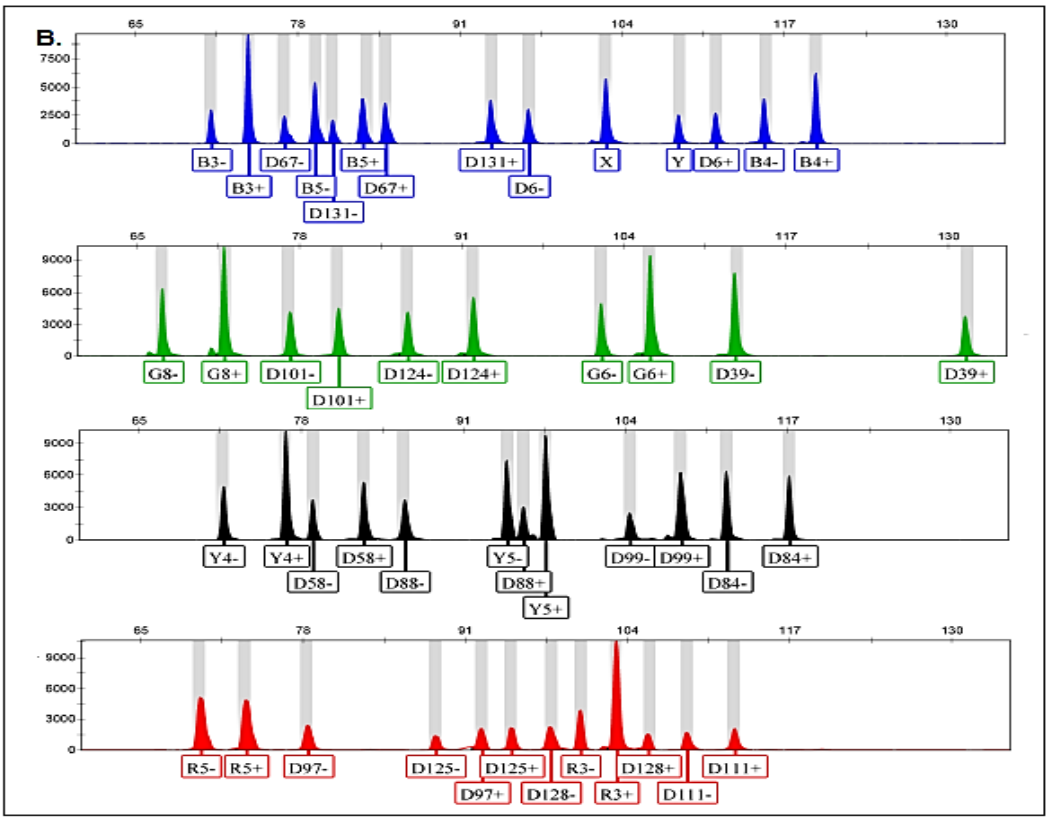


Figure 7-18: Electropherogram shows all 48 alleles present in the ladder on Mini-INDELs multiplex using (A) 0.5 ng and (B) 1.0 ng of DNA mixed.

7.4 Discussion

Chapter 7 describes the development and optimisation of a mutilplex mini-INDEL assay containing nine primer pairs re-designed from privies assay with reduced amplicon size. This assay was tested on Malay population samples, population genetic and forensic-related parameters calculated for the nine loci. The nine locus assay was then combined with a previously-developed 15 locus mini-INDEL assay, to produce a 24-locus mini-INDEL assay.

Table 7-15 Mini-INDELs multiplex were redesigned from larger loci in the DIPplex kit (including amelogenin) and addition of another 9 INDEL markers.

INDEL Markers	Qiagen Investigator™ DIPplex kit	In house Mini-Indel multiplex	INDEL Markers	Qiagen Investigator™ DIPplex kit	In house Mini-Indel multiplex
Amelogenin X	/	/	HLD124	/	
Amelogenin Y	/	/	HLD122	/	
HLD77	/		HLD125	/	/
HLD45	/		HLD64	/	
HLD131	/	/	HLD81	/	
HLD70	/		HLD136	/	
HLD6	/	/	HLD133	/	
HLD111	/	/	HLD97	/	/
HLD58	/	/	HLD40	/	
HLD56	/		HLD128	/	/
HLD118	/		HLD39	/	/
HLD92	/		HLD84	/	/
HLD93	/		B3		/
HLD99	/	/	B5		/
HLD88	/	/	B4		/
HLD101	/		G8		/
HLD67	/	/	G6		/
HLD83	/		Y4		/
HLD114	/		Y5		/
HLD48	/		R5		/
			R3		/

7.4.1 Primer design

The sequence data of all of Mini-INDELs which were available for different population groups from NCBI website. Then, Primer BLAST software was used to re-design primer sets. The self-complementarity of individual primers was assessed using the National Centre for Biotechnology Information (NCBI) Primer-BLAST (basic local alignment search tool) program (www.ncbi.nlm.nih.gov/Entrez/).

9 Mini-INDELs multiplex reactions were redesigned. The allele size ranges from 72 to 125 bp which presented in Table 7-1. The primer length of all the primers was less than 31 nucleotides and the G+C contents of each primer were between 25-60%. This multiplex amplified 9 loci. This included 3 FAM labelled primers sets (B3, B4 and B5), 2 NED labelled primers sets (G6 and G8), 2 VIC labelled primers sets (Y4 and Y5) and 2 PET labelled primers sets (R3 and R5). Below shows a flow diagram to develop a multiplex for the success of the reaction to be able to amplify all loci.

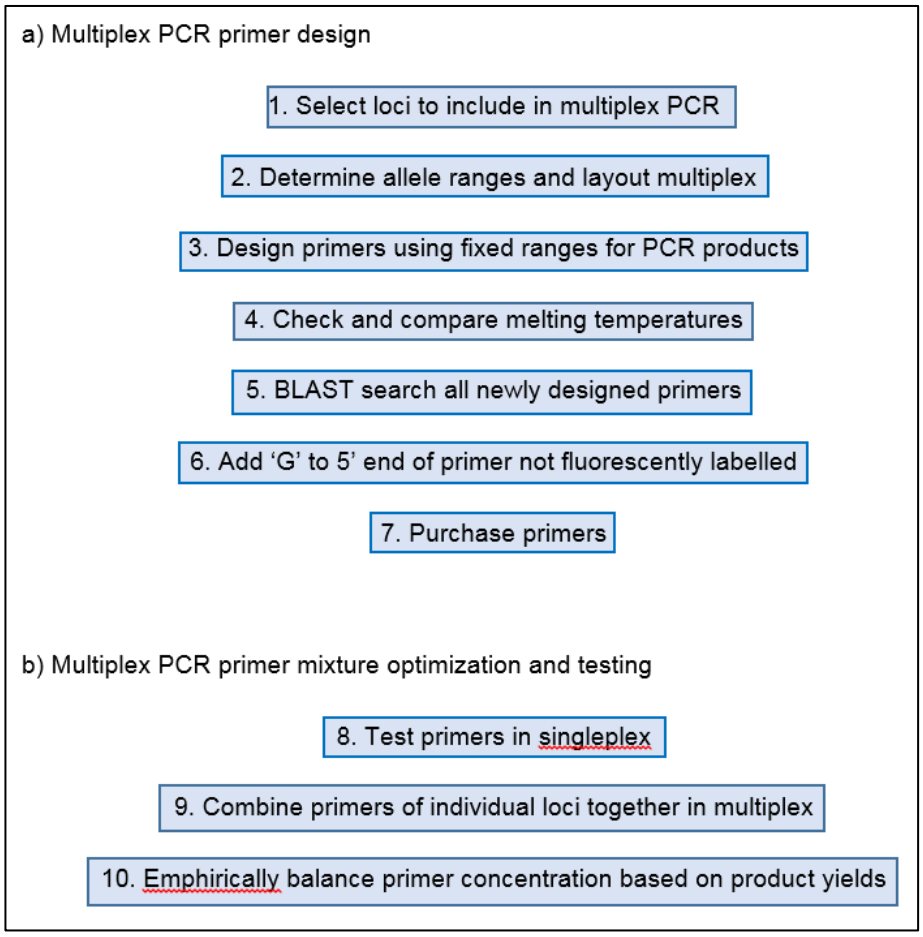


Figure 7-19 Flow diagram of a primer design-based approach in order to design multiplex reactions for the amplification loci adapted from (Schoske et al. 2003).

7.4.2 Optimisation of the multiplex

7.4.2.1 Singleplex

Primers for each locus were initially tested in a individually PCR reaction to evaluate the performance The criteria for primer failure are defines as those that produce profiles that exhibits incomplete adenylation, the presence of PCR artefacts, low signal, non-specific products or no PCR product at all (Zhang et al. 2015, Zhang et al. 2014). Once the successful primers at each locus were determined, those were equally combined together for a primer mix of 0.3 µM. Based on the results of genotyping profiles, the optimisation of each primer's concentration in the final primer mix was performed. Then the primers were tested also with 0.5 uM and 1.0 uM, and final concentration of each optimized primer are listed in Table.7-4.

7.4.2.2 Mini-INDEL multiplexes

(Wiegand, Kleiber 2001) had described that challenged samples with low DNA template could type successfully by using reduced primers that were close to target STR repeat compared to the established sequences that generated longer amplicons for the same loci. This same principle is applied in this research, to develop a multiplex PCR assay based on the INDELs which can type short amplicons for the analysis of highly degraded DNA or low template samples.

After the optimisation each primer pairs, the multiplex of 9 Mini-INDELs was successfully developed in a single tube reaction. The multiplex was optimised on amplification using different sets of PCR volumes, primer concentration, primer mix concentrations, DNA template, annealing temperature and extension time. It was found that this multiplex worked efficiently using primer concentrations of Set E with PCR final volume of 12.5 µl (Set 5). The optimum parameters were found to be 0.5 µM of primer mix concentration, 58 °C of annealing temperature and 30 minutes of extension time. The 9 Mini-INDEL investigated has demonstrated the ability to produce consistent and accurate genotype profiles for lower concentrations of template DNA, 0.3 ng, which suited for the analysis of degraded DNA (Song F 2015). Like the DIPplex® kit, 0.5 ng is the optimum DNA input which produced high sensitivity and peak height above 50 threshold value.

Figure 7-4 presented a typical electropherogram obtained with 9 Mini-INDELs assay. All 9 loci amplified well and the peak heights were reasonably balanced except for B3, Y4 an R3. More rigorous optimization could balance the amplification products for markers

B3, Y4 an R3 together with other markers in this multiplex reaction. However, because all the 9 loci consistently amplified, this multiplex can have considered acceptable as a means to generate population data. A comprehensive population study was therefore conducted on 100 Malay reference samples to determined allele frequency and various forensic parameters.

During the combination of 9 Mini-INDELs and 15 Mini-INDELs, there have been two problems where primer-primer interactions appear to have occurred. The first occurred in D67+ with B4- and D111+ with R3+., where the primers produced overlap. The second is formation of split peak in Y5- and D99 due to the distance between these primers were only 1 bp. To overcome this problem, new reverse primers were redesigned, resynthesized and optimised again for B4, Y5 and R3. There a few strategies to modify adenylation which include to redesigned reverse primer by adding a tail of 5 bases (TDAAT) or increased $MgCl_2$, lower annealing temperature, increase longer extension time and add more PCR cycles (Ballard 1997, Ballard, Scott & Rosatog 2002). All three alleles were separated and showed two distinctive peaks.

Alleles are designated where the difference in band size are within 0.5 bp and the absolute shift is no more than 0.5 bp. However, the mobility of the strands can be strongly influnces in denaturing conditions by a slight difference in (Adenine-Cytosine) A-C or Thymine-Cytosine (T-C) proportion (Carracedo, et al., 1988). The shifting from the expected size also due to the purification. Even though the primers ordered were HPLC purified, it is more or less to 85% purity according to most of commercial suppliers, which is good enough for PCR applications. Moreover, for labelled probes used, purification needed to ensure removal of free uncouple dyes, yet the difference of de-salted and HPLC purified primers are not significant. The reverse primers usually are not HPLC purified; where it was prepared with standard desalting.

7.4.2.3 Forensic parameters using the 9 Mini-INDELs

The forensic efficiency of 9 Mini-INDELs was evaluated by calculating combined match probability CPM, combined power of discrimination CPD and combined probability of exclusion CPE for the Malay subpopulations and compared with those amplified with DIPplex® kit, STRs and SNPs, also with Pakistani populations using the 14 Mini-INDELs assay (Table 7-16). As expected, all values of CPM, CPD and CPE are lower for 9 Mini-INDELs were lower but still they were significant.

The polymorphic nature makes them the markers of choice in characterization and genetic diversity studies. The high PIC values of the selected markers confirm their usefulness for the genetic polymorphism studies and linkage mapping programs in human as well. However, with a larger marker set; e.g combination of both Mini-INDELs (15 loci and 9 loci Mini-INDELs) will be necessary to further produce higher values in forensic parameters for individual identification also to characterized individuals within population groups on global level (e.g. Afirica, Europe and Asia) (Zaumsegel, Rothschild & Schneider 2013).

Table 7-16: Forensic parameters calculated for 9 Mini-INDELs from Malay subpopulation group. CPM represents combined match probability, CPD represents combined power of discrimination and CPE represents combined probability of exclusion for different kit used on the Malaysian population.

Forensic multiplex	Loci	Combined match probability (CPM)	Combined power of discrimination (CPD)	Combined probability of exclusion (CPE)	Ethnics in Malaysian Population
Mini-INDEL (in this study)	9	3.06×10^{-4}	99.243618%	97.64%	Malay
Mini-INDEL[1]	14	7.65×10^{-7}	99.9998%	96.34%	Pakistani
Investigator DIPplex® kit (in this study)	30	7.20×10^{-12}	99.9999%	99.99%	Malay
AmpFISTR Identifiler[3]	15	2.6×10^{-17}	99.9999%	99.99%	Malay
SNP[4]	52	4.10×10^{-19}	>99.999%	99.99%	Malay

1 (Majid Bashir PhD thesis; 2 (Lim et al. 2001); 2 (Seah et al. 2003); 3 Sharizah Alimat PhD thesis (2014).

7.4.3 Development of Allelic ladder

The allelic ladder was successfully developed by combining all the heterozygous alleles for all 24 Mini-INDEL markers; contained both the insertion deletion alleles of every marker in the panel. It was not possible to achieve a complete balance between all markers, as some imbalances consistently re-appeared after re-amplification of the ladder master mix. The remaining imbalances however had no impact on the consistency of the automated allele calling performed by the GeneMapper IDX software (Applied Biosystems).

When allelic ladder has developed, accuracy can be determined by referring to the ability of the method to obtain a correct size and genotype (Krenke et al. 2002, Daniels, Hall & Ballantyne 2004). The standard DNA sample Human Genome Control (Promega) was selected for this study. All DNA samples were amplified and typed in duplicate for every analysis. Accuracy was expressed as the variation in size (bp) between the allele(s) of the sample compared to those same allele(s) within the allelic ladder of the same run. All sizes were determined using an internal size standard GeneScan LIZ 500. In addition, in the accuracy test alleles across all loci fell within ± 0.5 bp of the same allele in the allelic ladder based on the bins created earlier. ABI Prism DNA Genotyper Analysis Software V2.1 allows for ± 0.5 bp variation between the unknown fragments and the fragments in the allelic ladder. In each set of analysis conducted, the standard sample, Human Genome DNA (Promega) fell within the ± 0.5 bp variation allowed. Furthermore, the standard sample, Promega Human Genome, did not show the present of variant allele.

7.5 Conclusion

The 9 loci Mini-INDELs were successfully amplified with a good peak height and expected product sizes. However, it is impractical to apply as supplementary or stand-alone assays for human identification since the discrimination power is still low. However, with the construction of the New Mini-INDELs multiplex (9 & 15 Mini-INDELs) will not only revealed potential advantages in the analysis of degraded samples, but also improving the discrimination power. These demonstrate the potential used of Mini-INDEL-plex as a complementary tool to INDELs and STRs in the analysis of forensic samples.

CHAPTER 8

CONCLUSION AND FUTURE WORK

8.1 General discussion

Today, forensic DNA analyst continue to process an increasingly large number of non-standard, complex evidentiary samples that contains low amount of DNA submitted by their law enforcement agencies. Because of their demand, samples such as touch DNA, mixed samples and samples with low amounts of biological material have become more frequently processed requiring more expertise and new tool to obtain profiles. For example, low quantity and/or poor quality of DNA are often recovered from those types of samples and contain PCR inhibitors that reduce the probability of obtaining results. In addition, like other forensic laboratories in tropical climates, many of the crime samples received are less than ideal. due to the factors like environmental exposure to heat, sun, and humidity for days or even months.

The main aim of this project are to evaluate whether the new indentification tool, the Insertion/Deletion Polymorphisms (INDELs) is suitable for forensic particularly for Malaysian samples where often in a poor quality due to expose in tropical environment; high humidity and heat. Following this, is to develop a new multiplex (mini INDEL) in order to analyse highly degraded samples. The second aim of this project was to develop a new multiplex (Mini 4-plex & IACs) that can be used to assess the quality and quantity of the DNA extracts in shorter fragments (50 bp, 70 bp, 90 bp, 112 bp, 154 bp and 170 bp). In the previous PhD research, the 4-plex multiplex without the IACs was developed for the degradation study in tissue samples (Nazir et al. 2012). But, problems arise when inhibited samples were being analysed using this 4-plex multiplex. This is because highly degraded DNA samples produced either partial DNA profiles or negative results and similar results also produced from inhibited samples which could be mistakenly assumed to be degraded DNA (Kontanis & Reed 2006). Thus two IACs were included to detect the presence of PCR inhibitors.

Chapter 3 presents about the performance of the Investigator DIPplex® kit (Qiagen). While there is no forensic database currently available for use with this kit, the INDEL profiles could be used to supplement STR profiles, although studies to establish the

linkage effects of STR loci and nearby INDELs would be required. Additional information from INDELs would be especially useful on occasions where only a partial STR profile is obtained. Analysis bi-allelic insertion/deletion polymorphisms can be used as stand-alone assays or supplemental data in cases when only partial STR profiles are generated from challenged and degraded samples.

The chapter shows that the kit is sensitive to relatively low levels of input DNA, compares the effectiveness of two DNA extraction methods in relation to the successful amplification of the kit markers, examines electropherogram features to determine the quality of INDEL profiles, consider the effect of DNA degradation on the production of INDEL profiles and examines the performance of the kit on samples from five Malaysian populations and on casework/bone samples, in comparison to STR analysis.

From the evaluation, the Qiagen Investigator DIPplex® is a very straightforward and unambiguous approach for the typing of degraded DNA, with particular effective application for identification and crime cases.

In **Chapter 4** reported that All INDEL loci are polymorphic in the population and the combined match probability (CMP) are 1 in 5.25×10^{-12} for Malay, 1 in 2.70×10^{-11} for Chinese, 1 in 1.79×10^{-12} for Indian, 1 in 4.87×10^{-11} for Iban, and 1 in 7.6×10^{-11} for Bidayuh. The combined power of discrimination is greater than 99.9999% and the combined power of exclusion are all 99.99%.

None of the tested 30 loci showed significant departure for HWE (after Bonferroni test), therefore they are statistically independent. The average F_{IS} value for Malaysian population has been calculated as 0.032 indicated low level of inbreeding within the population and that the population is in random mating with a high level of heterozygosity.

These results demonstrate the potential use of INDEL for identification and as a complementary tool to STRs in the analysis of forensic samples. It is also an efficient alternative SNP typing in the studied population. This report of allele frequencies of INDELs would serve as reference database for individual identification in the Malaysian population in the future. This database also has value for other population genetics and diversity studies since INDELs are stable genetic markers for lineage-based analysis even without reference sample (Budowle, van Daal 2008).

In this study also, the INDEL markers were applied to the crime and bone samples, in order to evaluate the usefulness of these INDELs in handling problematic samples. The

results showed that the INDEL assay returns more genetic information for blood stains and bone samples, compared to the STR assay. Even the match probability of the STR is higher, INDELs still gives an acceptable value for forensic identification; e.g. linking different pieces of evidence or re-association of body parts in the case of human identification.

In general, the populations which were carried out in this study are closely related to be properly differentiated with small sets of 30 autosomal INDEL markers. The 46-plex INDEL set can be used in the future as there are designed for broad-scale continental comparisons. As a conclusion, the Qiagen Investigator® DIPplex kit are not well suited to study of geographically close populations because it was developed for forensic purposes.

The mini 4-plex & IACs PCR assay able to provide qualitative information about the level of degradation and the presence of PCR inhibitors. The appropriate use of IACs in samples containing PCR inhibitors can further increase the first-pass STR typing efficiency and minimize the number of samples that require re-analysis and/or time consuming interpretation of suboptimal typing results. Finally, the results from the Mini 4-plex assay can also be used to confirm the results of DNA typing, which may prevent unnecessary re-analysis and re-analysis of degraded or inhibited samples that have already yielded the maximum amount of information available. The detail of this study has been discussed **Chapter 5**. The key application of the method is the ability to triage samples prior to attempting to produce a full STR, mini-STR or INDEL profile.

Also, PCR inhibition can be detected through the use of an internal PCR control (IPC) during real-time quantification (Seo et al. 2012, Kontanis & Reed 2006). This technique could save more time and cost since less expensive reagents are involved while the Mini 4-plex & IACs multiplex has to be analyzed on a Genetic Analyzer to evaluate the quality of a sample which will caused more time and cost. Accurate quantification result also could be obtained from the real-time quantification, while the estimation of DNA quantification using this multiplex was very poor. As described in **Chapter 6**, a correlation graph was plotted based on the average peak heights of mini 4-plex and DNA concentrations from the serial diluted control DNA samples. This correlation graph has coefficient of determination (R2) of 0.9741 and can only be used when the sample concentrations are between 0.07 ng – 2.5 ng. Also, only 10% of tested samples produced relative standard deviations below 10% indicating inaccuracy of this correlation graph. However, there are limitations with real-time quantification approach where the amplicon

in the PCR reactions is typically short, and so does not necessarily reveal the full extent of PCR inhibition. But with the mini 4-plex & IACs results, full extend of the sample quality could be revealed including the DNA degradation and the presence of PCR inhibition.

Subsequently **Chapter 7** present about developing a simple multiplex assays to fill the gaps of the Mini-INDELs multiplex which was developed before. There are several crucial criteria that need to be carefully evaluated, especially primer-primer concentration and interactions concentrations used for PCR buffers and other components, thermal conditions and CE instruments used to analyse the samples.

Overall from this research, forensic scientist will be able to accurately investigate the quantity of highly degraded DNA present in forensic samples and simultaneously determine the most appropriate typing system for each samples. Indirectly it will increase the efficiency of the laboratory and more prudent use of limited samples. In addition, the approaches and methods (see Figure 8-1) of choice provide robust, sensitive and cost effective genotyping for human identification.

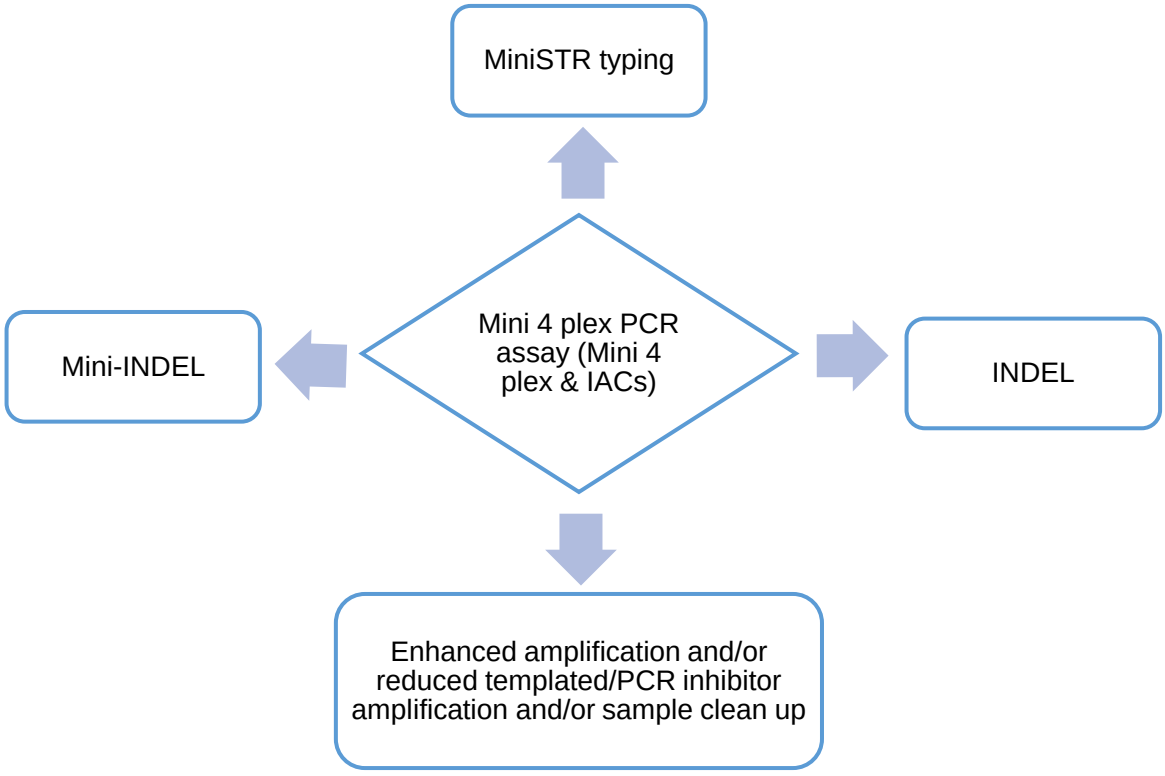


Figure 8-1: Flowchart shows the choice of DNA typing strategy after analyze with the Mini 4-plex PCR assay (edited from Hudlow R.W. et al 2008)

With this research, the assays could be applied to Malaysia lab and allow DNA to be processed efficiently, help to process more complex samples and lastly reduce backlogs.

8.2 Limitations

The strongest part of this work is the multiplex are designed specifically to overcome the issues especially in Malaysia. This will increase the efficiency of DNA evidence beside using in-house kits which could save cost, time and even samples. The highlight of this thesis is the development of the Mini 4-plex & IACs. This assay has the potential over commercial kits available especially in terms of cost and time. Beside the Mini 4-plex could be used in fully extent conditions.

Alternatively, the current forensic laboratory in Malaysia already have the infrastructure to genotype INDEL markers, as the workflow is identical to STR genotyping, and these markers also show an improvement over STR loci for shorter fragments; below 200 bp. However, the only issue need to be taken care is the imbalance of the peaks especially in Mini-INDEL. This is where the multiplex still need to be improved in order to obtain more quality profiles that provide consistent and corrective results with downstream

The analysis conducted using the 30 biallelic markers which are very limited for distinguishing population studies which are closely related. The 46-plex INDEL set can be used in the future as there are designed for broad-scale continental comparisons. As a conclusion, the Qiagen Investigator® DIPplex kit are not well suited to study of geographically close populations.

Due to limit of time, only populations samples and a few casework samples were analysed using both INDEL Investigator Kit and Mini-INDEL. It is good to see the effectiveness of the assays on difficult samples such as touch DNA, mixed stain samples and difficult paternity/kinship in comparison with the STR kits.

8.3 Future work

These results demonstrate the potential use of INDEL for identification and as a complementary tool to STRs in the analysis of forensic samples. It is also an efficient alternative SNP typing in the studied population. This report of allele frequencies of INDELs would serve as reference database for individual identification in the Malaysian population in the future. This database also has value for other population genetics and diversity studies since INDELs are stable genetic markers for lineage-based analysis

even without reference sample (Budowle, van Daal 2008); i.e murder case where the suspect's reference unable to obtain for comparison with crime scene samples.

As these marker types become more accessible to forensic laboratories, more study of their utility for low template DNA sample is needed; analysing touch DNA samples, semen stains and challenging samples which have been exposed to high temperature and humidity for a length of time.

This research has shed the light on unexplored area on forensic DNA in managing degraded samples. The studies carried out showed in practical terms the applicability of this multiplex (4-plex and IACs) assay within the DNA profiling framework, giving confidence in the results obtained particularly with respect to challenging samples such as degraded and inhibited samples. As future study can be conducted by adding primer to quantify total human male and to quantify human male DNA in the Mini 4-plex to assess samples from rape cases.

Further population studies on variation in individual admixture proportions will be interesting for Malaysian's multi-ethnic populations especially on Orang Asli by performing mitochondrial method. Although the basic mitochondrial DNA protocols and practices are similar, an overview of the current practices of forensic mtDNA analysis using the latest sequence technology NGS will provide and help to frame the path forward. In addition, the current interest that growing is establishing SNP tests that are able to predict hair, skin and eye colour variation in populations, along with the ancestry informative markers (AIM) (Li et al., 2008, Fondevila et al., 2013, Pereira et al., 2012) that associate the genotypes detected in an individual to their genetic ancestry, where in this case, genetic ancestry is a characteristic defined by broadly based continental population group.

REFERENCES

1000 Genomes Project Consortium 2012, "An integrated map of genetic variation from 1,092 human genomes", *Nature,* vol. 491, no. 7422, pp. 56-65.

Abu Al-Soud, W. 2000, *Optimisation of diagnostic PCR a study of PCR inhibitors in blood and sample pretreatment*, Department of Chemistry,Lund University, Lund, Sweden.

Adamowicz, M.S., Stasulli, D.M., Sobestanovich, E.M. & Bille, T.W. 2014, "Evaluation of methods to improve the extraction and recovery of DNA from cotton swabs for forensic analysis", *PloS one,* vol. 9, no. 12, pp. e116351.

Aghakhanian, F., Yunus, Y., Naidu, R., Jinam, T., Manica, A., Hoh, B.P. & Phipps, M.E. 2015, "Unravelling the genetic history of Negritos and indigenous populations of Southeast Asia", *Genome biology and evolution,* vol. 7, no. 5, pp. 1206-1215.

Akane, A., Matsubara, K., Nakamura, H., Takahashi, S.i. & 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,* vol. 39, pp. 362-362.

Akhteruzzaman, S., Das, S.A., Hosen, I. & Ferdous, A. 2013, "Genetic polymorphism of 30 InDel markers for forensic use in Bangladeshi population", *Forensic Science International: Genetics Supplement Series,* vol. 4, no. 1, pp. e348-e349.

Alaeddini, R. 2012, "Forensic implications of PCR inhibition—a review", *Forensic Science International: Genetics,* vol. 6, no. 3, pp. 297-305.

Alaeddini, R., Walsh, S.J. & Abbas, A. 2010, "Forensic implications of genetic analyses from degraded DNA—a review", *Forensic science international: genetics,* vol. 4, no. 3, pp. 148-157.

Al-Soud, W.A., Ouis, I.S., Li, D.Q., Ljungh, Å. & Wadström, T. 2005, "Characterization of the PCR inhibitory effect of bile to optimize real-time PCR detection of Helicobacter species", *FEMS Immunology & Medical Microbiology,* vol. 44, no. 2, pp. 177-182.

Al-Soud, W.A. & Radstrom, P. 2001, "Purification and characterization of PCR-inhibitory components in blood cells", *Journal of clinical microbiology,* vol. 39, no. 2, pp. 485-493.

Alvarez, M.F. 2011, "Forensic Performance of Short Amplicon Insertion-Deletion (InDel) markers", *22nd International Symposium on Human Identification*, eds. R. Pereira, L. Gusmao, C. Phillips, et al, .

Amorim, A. & Pereira, L. 2005, "Pros and cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs", *Forensic science international,* vol. 150, no. 1, pp. 17-21.

Andaya, L.Y. 2001, "The search for the “origins” of Melayu", *Journal of Southeast Asian Study, (2001), pp. 1–16,* vol. 32, pp. 1–16.

Arezi, B., Xing, W., Sorge, J.A. & Hogrefe, H.H. 2003, "Amplification efficiency of thermostable DNA polymerases", *Analytical Biochemistry,* vol. 321, no. 2, pp. 226-235.

Asamura, H., Fujimori, S., Ota, M. & Fukushima, H. 2007, "MiniSTR multiplex systems based on non-CODIS loci for analysis of degraded DNA samples", *Forensic science international,* vol. 173, no. 1, pp. 7-15.

Aşıcıoğlu, F., Oguz, S.F. & Ozbek, U. 2004, "Mutation rate at commonly used forensic STR loci: Paternity testing experience", *Disease markers,* vol. 20, no. 6, pp. 313-315.

Ballantyne, J. 2006, *Assessment and in vitro repair of damaged DNA templates,* National Ctr for Forensic Science, United States of America.

Ballard, L.W. 1997, "High-throughput genetic analysis in the core facility", *ABRF News,* vol. 8, pp. 22-28.

Ballard, L.W., Scott, P. & Rosatog, C. 2002, "Strategies for Genotyping: Effectiveness of Tailing Primers to Increase Accuracy in Short Tandem Repeat Determinations", *Journal of Biomolecular Techniques,* vol. 13, no. 1, pp. 20-29.

Bandelt, H.J. & Salas, A. 2012, "Current next generation sequencing technology may not meet forensic standards", *Forensic Science International: Genetics,* vol. 6, no. 1, pp. 143-145.

Barbaro, A., Cormaci, P. & Agostino, A. 2009, "Validation of PrepFiler™ forensic DNA extraction kit (Applied Biosystems)", *Forensic Science International: Genetics Supplement Series,* vol. 2, no. 1, pp. 176-177.

Barker, G., Barton, H., Bird, M., Daly, P., Datan, I., Dykes, A., Farr, L., Gilbertson, D., Harrisson, B. & Hunt, C. 2007, "The ‘human revolution’in lowland tropical Southeast Asia: the antiquity and behavior of anatomically modern humans at Niah Cave (Sarawak, Borneo)", *Journal of human evolution,* vol. 52, no. 3, pp. 243-261.

Bashir, M. & Hassan, N.H. 2016, "Analysis of 30 Biallelic INDEL Markers Using the Investigator DIPplex((R)) Kit", *Methods in molecular biology (Clifton, N.J.),* vol. 1420, pp. 135-142.

Bastos-Rodrigues, L., Pimenta, J.R. & Pena, S.D.J. 2006, "The genetic structure of human populations studied through short insertion-deletion polymorphisms", *Annals of Human Genetics,* vol. 70, no. 5, pp. 658-665.

Bej, A.K., Mahbubani, M.H., Dicesare, J.L. & Atlas, R.M. 1991, "Polymerase chain reaction-gene probe detection of microorganisms by using filter-concentrated samples", *Applied and Environmental Microbiology,* vol. 57, no. 12, pp. 3529-3534.

Belec, L., Authier, J., Eliezer V., M.C., Piédouillet, C., Mohamed, A.S. & Gherardi, R.K. 1998, "Myoglobin as a polymerase chain reaction (PCR) inhibitor: a limitation for PCR from skeletal muscle tissue avoided by the use of Thermus thermophilus polymerase", *Muscle & nerve,* vol. 21, no. 8, pp. 1064-1067.

Bender, K., Farfán, J. & Schneider, P.M. 2004, "Preparation of degraded human DNA under controlled conditions", *Forensic science international,* vol. 139, no. 2, pp. 135-140.

Benson, D.A., Boguski, M.S., Lipman, D.J., Ostell, J., Ouellette, B.F., Rapp, B.A. & Wheeler, D.L. 1999, "GenBank", *Nucleic acids research,* vol. 27, no. 1, pp. 12-17.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. & Wheeler, D.L. 2005, "GenBank", *Nucleic acids research,* vol. 33, no. Database issue, pp. D34-8.

Bergallo, M., Costa, C., Gribaudo, G., Tarallo, S., Baro, S., Ponzi, A.N. & Cavallo, R. 2006, "Evaluation of six methods for extraction and purification of viral DNA from urine and serum samples", *New Microbiologica,* vol. 29, no. 2, pp. 111-119.

Bessetti, J. 2007, "An introduction to PCR inhibitors", *J.Microbiol.Meth,* vol. 28, pp. 159-167.

Bland, J.M. & Altman, D.G. 1995, "Multiple significance tests: the Bonferroni method", *BMJ (Clinical research ed.),* vol. 310, no. 6973, pp. 170.

Brevnov, M.G., Pawar, H.S., Mundt, J., Calandro, L.M., Furtado, M.R. & Shewale, J.G. 2009, "Developmental validation of the PrepFiler™ forensic DNA extraction kit for extraction of genomic DNA from biological samples", *Journal of forensic sciences,* vol. 54, no. 3, pp. 599-607.

Briggs, A.W., Good, J.M., Green, R.E., Krause, J., Maricic, T., Stenzel, U., Lalueza-Fox, C., Rudan, P., Brajkovic, D., Kucan, Z., Gusic, I., Schmitz, R., Doronichev, V.B., Golovanova, L.V., de la Rasilla, M., Fortea, J., Rosas, A. & Paabo, S. 2009, "Targeted retrieval and analysis of five Neandertal mtDNA genomes", *Science (New York, N.Y.),* vol. 325, no. 5938, pp. 318-321.

Bright, J.-, Curran, J.M. & Buckleton, J.S. 2013, "Relatedness calculations for linked loci incorporating subpopulation effects", *Forensic Science International: Genetics,* vol. 7, no. 3, pp. 380-383.

Bright, J.-, Hopwood, A., Curran, J.M. & Buckleton, J.S. 2014, "A Guide to Forensic DNA Interpretation and Linkage", *Promega Corporation Web site,* [Online], , pp. 21/09/2016. Available from: http://www.promega.com/ resources/profiles-in-dna/2014/a-guide-to-forensic-dna-interpretation-and-linkage/.

Bright, J.-, Taylor, D., Curran, J.M. & Buckleton, J.S. 2013, "Degradation of forensic DNA profiles", *Australian Journal of Forensic Sciences,* vol. 45, no. 4, pp. 445-449.

Bright, J.-, Turkington, J. & Buckleton, J. 2010, "Examination of the variability in mixed DNA profile parameters for the Identifiler™ multiplex", *Forensic science international: genetics,* vol. 4, no. 2, pp. 111-114.

Buckleton, J. & Triggs, C. 2006, "The effect of linkage on the calculation of DNA match probabilities for siblings and half siblings", *Forensic science international,* vol. 160, no. 2, pp. 193-199.

Budowle, B. & van Daal, A. 2008, "Forensically relevant SNP classes", *BioTechniques,* vol. 44, no. 5, pp. 603-+.

Bulbeck, D. 2011, "10 Biological and cultural evolution in the population and culture history of Homo sapiens in Malaya", *Dynamics of human diversity,* , pp. 207.

Burkhart, C.A., Norris, M.D. & Haber, M. 2002, "A simple method for the isolation of genomic DNA from mouse tail free of real-time PCR inhibitors", *Journal of Biochemical and Biophysical Methods,* vol. 52, no. 2, pp. 145-149.

Butler, J.M. 2014, *Advances Topics in Forensic DNA Typing: Interpretation,* Academic Press.

Butler, J.M. 2011, *Advanced Topics in Forensic DNA Typing: Methodology: Methodology,* Academic Press.

Butler, J.M. 2007, "Short tandem repeat typing technologies used in human identity testing", vol. 43, no. 4, pp. 2-5.

Butler, J.M. 2005, *Forensic DNA typing. Biology, Technology, and Genetics of STR markers,* 2nd edition edn, Elsevier Academic Press, USA.

Butler, J.M., Coble, M.D. & Vallone, P.M. 2007, "STRs vs. SNPs: thoughts on the future of forensic DNA testing", *Forensic science, medicine, and pathology,* vol. 3, no. 3, pp. 200-205.

Butler, J.M. & Hill, C.R. 2010, "Scientific issues with analysis of low amounts of DNA", *Profiles DNA,* [Online], vol. 13, no. Promega Corporation Web site, pp. 17/09/2016. Available from: http://www.promega.co.uk/resources/profiles-in-dna/2010/scientific-issues-with-analysis-of-low-amounts-of-dna/.

Butler, J.M., Shen, Y. & McCord, B.R. 2003, "The development of reduced size STR amplicons as tools for analysis of degraded DNA", *Journal of forensic sciences,* vol. 48, no. 5, pp. 1054-1064.

Cardoso, S., Sevillano, R., Gamarra, D., Santurtún, A., Martínez-Jarreta, B. and de Pancorbo, M.M., 2017. Population genetic data of 38 insertion-deletion markers in six populations of the northern fringe of the Iberian Peninsula. *Forensic Science International: Genetics*, *27*, pp.175-179.

Carvalho, A. & Pinheiro, M.F. 2013, "Population data of 30 insertion/delection polymorphisms from a sample taken in the North of Portugal", *International journal of legal medicine,* , pp. 1-3.

Chang, Y.M., Perumal, R., Keat, P.Y. & Kuehn, D.L.C. 2007, "Haplotype diversity of 16 Y-chromosomal STRs in three main ethnic populations (Malays, Chinese and Indians) in Malaysia", *Forensic science international,* vol. 167, no. 1, pp. 70-76.

Chang, Y.M., Swaran, Y., Phoon, Y.K., Sothirasan, K., Sim, H.T., Lim, K.B. & Kuehn, D. 2009, "Haplotype diversity of 17 Y-chromosomal STRs in three native Sarawak populations (Iban, Bidayuh and Melanau) in East Malaysia", *Forensic Science International: Genetics,* vol. 3, no. 3, pp. e77-e80.

Coble, M.D. & Butler, J.M. 2005, "Characterization of new miniSTR loci to aid analysis of degraded DNA", *Journal of forensic sciences,* vol. 50, no. 1, pp. 43-53.

Collins, P.J., Hennessy, L.K., Leibelt, C.S., Roby, R.K., Reeder, D.J. & Foxall, P.A. 2004, "Developmental validation of a single-tube amplification of the 13 CODIS STR loci, D2S1338, D19S433, and amelogenin: the AmpFlSTR Identifiler PCR Amplification Kit", *Journal of forensic sciences,* vol. 49, no. 6, pp. 1265-1277.

Cotton, E., Allsop, R., Guest, J., Frazier, R., Koumi, P., Callow, I., Seager, A. & Sparkes, R. 2000, "Validation of the AMPFlSTR® SGM Plus™ system for use in forensic casework", *Forensic science international,* vol. 112, no. 2, pp. 151-161.

da Costa Francez., P. A., Ribeiro-Rodrigues, E.M. & 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,* vol. 6, no. 1, pp. 132-135.

da Silva, C.V., Matos, S., Costa, H.A., Morais, P., Dos Santos, R.M., Espinheira, R., Santos, J.C. & Amorim, A. 2013, "Genetic portrait of south Portugal population with InDel markers", *Forensic science international.Genetics,* vol. 7, no. 4, pp. e101-3.

Daniels, D.L., Hall, A.M. & Ballantyne, J. 2004, "SWGDAM developmental validation of a 19-locus Y-STR system for forensic casework", *Journal of Forensic Science,* vol. 49, no. 4, pp. 1-16.

Dannemiller, K.C., Weschler, C.J. & Peccia, J. 2016, "Fungal and bacterial growth in floor dust at elevated relative humidity levels", *Indoor air,*

Dawson, E., Abecasis, G.R., Bumpstead, S., Chen, Y., Hunt, S., Beare, D.M., Pabial, J., Dibling, T., Tinsley, E., Kirby, S. and Carter, D., 2002. A first-generation linkage disequilibrium map of human chromosome 22. *Nature*, *418*(6897), pp.544-548..

de Camargo, M.M. & Nahum, L.A. 2005, "Adapting to a changing world: RAG genomics and evolution", *Human Genomics,* vol. 2, no. 2, pp. 1-6.

Deagle, B.E., Eveson, J.P. & Jarman, S.N. 2006, "Quantification of damage in DNA recovered from highly degraded samples–a case study on DNA in faeces", *Frontiers in Zoology,* vol. 3, no. 1, pp. 1.

Deng, L., Hoh, B.P., Lu, D., Fu, R., Phipps, M.E., Li, S., Hatin, W.I., Ismail, E., Mokhtar, S.S. & Jin, L. 2014, "The population genomic landscape of human genetic structure, admixture history and local adaptation in Peninsular Malaysia", *Human genetics,* vol. 133, no. 9, pp. 1169-1185.

Deng, L., Hoh, B.P., Lu, D., Saw, W.Y., Twee-Hee Ong, R., Kasturiratne, A., de Silva, H.J., Zilfalil, B.A., Kato, N., Wickremasinghe, A.R., Teo, Y.Y. & Xu, S. 2015,

"Dissecting the genetic structure and admixture of four geographical Malay populations", *Scientific reports,* vol. 5, pp. 14375.

Department of Orang Asli Development Malaysia 2013, *Sub-tribe/Ethnic.* Ministry of Rural and Regional Development, Malaysia.

Department of Statistics, M. 2010, *Population distribution and basic demographic characteristics,* Putrajaya, Malaysia:.

Dixon, L., Dobbins, A., Pulker, H., Butler, J., Vallone, P.M., Coble, M., Parson, W., Berger, B., Grubwieser, P. & Mogensen, H.S. 2006, "Analysis of artificially degraded DNA using STRs and SNPs—results of a collaborative European (EDNAP) exercise", *Forensic science international,* vol. 164, no. 1, pp. 33-44.

Dong, C.N., Yang, Y.D., Li, S.J., Yang, Y.R., Zhang, X.J., Fang, X.D., Yan, J.W. & Cong, B. 2016, "Whole genome nucleosome sequencing identifies novel types of forensic markers in degraded DNA samples", *Scientific reports,* vol. 6, pp. 26101.

Eckhart, L., Bach, J., Ban, J. & Tschachler, E. 2000, "Melanin binds reversibly to thermostable DNA polymerase and inhibits its activity", *Biochemical and biophysical research communications,* vol. 271, no. 3, pp. 726-730.

El Bali, L., Diman, A., Bernard, A., Roosens, N.H. & De Keersmaecker, S.C. 2014, "Comparative study of seven commercial kits for human DNA extraction from urine samples suitable for DNA biomarker-based public health studies", *Journal of biomolecular techniques : JBT,* vol. 25, no. 4, pp. 96-110.

Excoffier, L. & 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,* vol. 10, no. 3, pp. 564-567.

Excoffier, L., Laval, G. & Schneider, S. 2007, "Arlequin (version 3.0): an integrated software package for population genetics data analysis", *Evolutionary bioinformatics online,* vol. 1, pp. 47-50.

Fachmann, M.S.R., Josefsen, M.H., Hoorfar, J., Nielsen, M.T. & Löfström, C. 2015, "Cost-effective optimization of real-time PCR-based detection of Campylobacter and Salmonella with inhibitor tolerant DNA polymerases", *Journal of applied microbiology,* vol. 119, no. 5, pp. 1391-1402.

Feine, I., Shpitzen, M., Roth, J. & Gafny, R. 2016, "A novel cell culture model as a tool for forensic biology experiments and validations", *Forensic Science International: Genetics,* vol. 24, pp. 114-119.

Fondevila, M., Phillips, C., Santos, C., Pereira, R., Gusmao, L., Carracedo, A., Butler, J.M., Lareu, M.V. & Vallone, P.M. 2012, "Forensic performance of two insertion-deletion marker assays", *International journal of legal medicine,* vol. 126, no. 5, pp. 725-737.

Fondevila, M., Phillips, C., Naverán, N., Cerezo, M., Rodríguez, A., Calvo, R., Fernández, L., Carracedo, Á. & Lareu, M. 2008, "Challenging DNA: assessment of

a range of genotyping approaches for highly degraded forensic samples", *Forensic Science International: Genetics Supplement Series,* vol. 1, no. 1, pp. 26-28.

Fordyce, S.L., Mogensen, H.S., Børsting, C., Lagacé, R.E., Chang, C.W., Rajagopalan, N. & Morling, N. 2015, "Second-generation sequencing of forensic STRs using the Ion Torrent™ HID STR 10-plex and the Ion PGM™", *Forensic Science International: Genetics,* vol. 14, pp. 132-140.

Forster, L., Thomson, J. & Kutranov, S. 2008, "Direct comparison of post-28-cycle PCR purification and modified capillary electrophoresis methods with the 34-cycle “low copy number”(LCN) method for analysis of trace forensic DNA samples", *Forensic Science International: Genetics,* vol. 2, no. 4, pp. 318-328.

Fredericks, J.D., Bennett, P., Williams, A. & Rogers, K.D. 2012, "FTIR spectroscopy: A new diagnostic tool to aid DNA analysis from heated bone", *Forensic Science International: Genetics,* vol. 6, no. 3, pp. 375-380.

Freitas, N.S.C., Resque, R.L., Ribeiro-Rodrigues, E.M., Guerreiro, J.F., Santos, N.P.C., Ribeiro-dos-Santos, A. & Santos, S. 2010, "X-linked insertion/deletion polymorphisms: forensic applications of a 33-markers panel", *International journal of legal medicine,* vol. 124, no. 6, pp. 589-593.

Friis, S.L., Borsting, C., Rockenbauer, E., Poulsen, L., Fredslund, S.F., Tomas, C. & Morling, N. 2012, "Typing of 30 insertion/deletions in Danes using the first commercial indel kit-Mentype (R) DIPplex", *Forensic Science International-Genetics,* vol. 6, no. 2, pp. E72-E74.

Gao, T., Yun, L., He, W., Gu, Y. & Hou, Y. 2015, "The application of multi-InDel as supplementary in paternity cases with STR mutation", *Forensic Science International: Genetics Supplement Series,* vol. 5, pp. e218-e219.

Garvin, M.R., Saitoh, K. & Gharrett, J. 2010, "Appicaltion of single nucleotide polymorphisms to non-model species: a technical review", *Molecular Ecology Resources,* vol. 10, pp. 915-934.

Geada, H., Viriato, L., Vieira-Silva, C., Cruz, C., Lucas, I., Ribeiro, T. & Espinheira, R. 2003, "STR mutations in paternity investigations: A study of 1-year consecutive cases", *Progress in Forensic Genetics 9,* vol. 1239, pp. 657-660.

Geng, T. & Mathies, R.A. 2015, "Minimizing inhibition of PCR-STR typing using digital agarose droplet microfluidics", *Forensic Science International: Genetics,* vol. 14, pp. 203-209.

Gentra® Puregene® Handbook June 2011, *Qiagen,* Third edn, © 2007–2011 QIAGEN.

Gettings, K.B., Kiesler, K.M. and Vallone, P.M 2015, "Performance of a next generation sequencing SNP assay on degraded DNA", Forensic Science International: Genetics, vol. 19, pp 1-9

Ghosh, R., Paul, S., Ghosh, S.K. & Roy, A. 2009, "An improved method of DNA isolation suitable for PCR-based detection of begomoviruses from jute and other mucilaginous plants", *Journal of virological methods,* vol. 159, no. 1, pp. 34-39.

Gill, P., Curran, J. & Elliot, K. 2005, "A graphical simulation model of the entire DNA process associated with the analysis of short tandem repeat loci", *Nucleic acids research,* vol. 33, no. 2, pp. 632-643.

Gomes, C., Magalhães, M., Alves, C., Amorim, A., Pinto, N. & Gusmão, L. 2012, "Comparative evaluation of alternative batteries of genetic markers to complement autosomal STRs in kinship investigations: autosomal indels vs. X-chromosome STRs", *International journal of legal medicine,* vol. 126, no. 6, pp. 917-921.

Gonzalez, A., Grimes, R., Walsh, E.J., Dalton, T. & Davies, M. 2007, "Interaction of quantitative PCR components with polymeric surfaces", *Biomedical Microdevices,* vol. 9, no. 2, pp. 261-266.

Goodwin, S., McPherson, J.D. & McCombie, W.R. 2016, "Coming of age: ten years of next-generation sequencing technologies", *Nature Reviews Genetics,* vol. 17, no. 6, pp. 333-351.

Goodwin, W., Linacre, A. & Hadi, S. 2011, *An introduction to forensic genetics,* John Wiley & Sons.

Gudnason, H., Dufva, M., Bang, D.D. & Wolff, A. 2007, "Comparison of multiple DNA dyes for real-time PCR: effects of dye concentration and sequence composition on DNA amplification and melting temperature", *Nucleic acids research,* vol. 35, no. 19, pp. e127.

Harismendy, O., Ng, P.C., Strausberg, R.L., Wang, X., Stockwell, T.B., Beeson, K.Y., Schork, N.J., Murray, S.S., Topol, E.J. & Levy, S. 2009, "Evaluation of next generation sequencing platforms for population targeted sequencing studies", *Genome biology,* vol. 10, no. 3, pp. 1.

Hatin, W.I., Zahri, M.K., Xu, S., Jin, L., Tan, S.G., Rizman, I.M., Zilfalil, A. & HUGO Pan-Asian SNP Consortium 2011, "Population genetic structure of peninsular Malaysia Malay sub-ethnic groups", *PloS one,* vol. 6, no. 4, pp. e18312.

Hedman, J., Nordgaard, A., Dufva, C., Rasmusson, B., Ansell, R. & Rådström, P. 2010, "Synergy between DNA polymerases increases polymerase chain reaction inhibitor tolerance in forensic DNA analysis", *Analytical Biochemistry,* vol. 405, no. 2, pp. 192-200.

Hedman, J., Nordgaard, A., Rasmusson, B., Ansell, R. & Rådström, P. 2009, "Improved forensic DNA analysis through the use of alternative DNA polymerases and statistical modeling of DNA profiles", *BioTechniques,* vol. 47, no. 5, pp. 951-958.

Hefke, G., Davison, S. & D'Amato, M.E. 2015, "Forensic performance of Investigator DIPplex indels genotyping kit in native, immigrant, and admixed populations in South Africa", *Electrophoresis,* vol. 36, no. 24, pp. 3018-3025.

Herrmann, B. & Henke, W. 1999, "DNA preservation: a microsatellite-DNA study on ancient skeletal remains", *Electrophoresis,* vol. 20, pp. 1722-1728.

Hildebrand, C.E., David, C., Torney, C. & Wagner, P. 1994, "Informativeness of polymorphic DNA markers", *The Human Genome Project: deciphering the blueprint of heredity.University Science Books, CA, USA,* , pp. 100-102.

Hill, C.R., Kline, M.C., Coble, M.D. & Butler, J.M. 2008, "Characterization of 26 miniSTR loci for improved analysis of degraded DNA samples", *Journal of forensic sciences,* vol. 53, no. 1, pp. 73-80.

Holt, C.L., Buoncristiani, M., Wallin, J.M., Nguyen, T., Lazaruk, K.D. & Walsh, P.S. 2002, "Twgdam Validation of Ampf_str•: PCR Amplification Kits for Forensic DNA Casework", *Journal of Forensic Science,* vol. 47, no. 1, pp. 66-96.

Hu, Q., Liu, Y., Yi, S. & Huang, D. 2015, "A comparison of four methods for PCR inhibitor removal", *Forensic Science International: Genetics,* vol. 16, pp. 94-97.

Huang, J., Luo, H., Wei, W. & Hou, Y. 2014, "A novel method for the analysis of 20 multi-Indel polymorphisms and its forensic application", *Electrophoresis,* vol. 35, no. 4, pp. 487-493.

Hughes-Stamm, S.R., Ashton, K.J. & Van Daal, A. 2011, "Assessment of DNA degradation and the genotyping success of highly degraded samples", *International journal of legal medicine,* vol. 125, no. 3, pp. 341-348.

HUGO Pan-Asian SNP Consortium 2009, "Mapping human genetic diversity in Asia", *Science,* vol. 326, no. 5959, pp. 1541-1545.

Ibarra, A., Martinez, M., Freire-Aradas, A., Fondevila, M., Carracedo, A., Porras, L. & Gusmão, L. 2013, "Using STR, MiniSTR and SNP markers to solve complex cases of kinship analysis", *Forensic Science International: Genetics Supplement Series,* vol. 4, no. 1, pp. e91-e92.

Imaizuml, K., Miyasaka, S. & Yoshino, M. 2004, "Quantitative analysis of amplifiable DNA in tissues exposed to various environments using competitive PCR assays", *Science & justice : journal of the Forensic Science Society,* vol. 44, no. 4, pp. 199-208.

Investigator DIPplex Handbook 2010, Qiagen.

Irwin, J., Just, R., Scheible, M. & Loreille, O. 2011, "Assessing the potential of next generation sequencing technologies for missing persons identification efforts", *Forensic Science International: Genetics Supplement Series,* vol. 3, no. 1, pp. e447-e448.

Iyavoo, S. 2014, *Development and application of a PCR multiplex to assess the quality and quantity of forensic DNA extracts*, University of Central Lancashire.

Jacobs, G. 2015, "Evaluation of insertion-deletion polymorphisms with the kit Qiagen Investigator® DIPplex for forensic application in South Africa", .

Jacobsen, C.S. & Rasmussen, O.F. 1992, "Development and application of a new method to extract bacterial DNA from soil based on separation of bacteria from soil

with cation-exchange resin", *Applied and Environmental Microbiology,* vol. 58, no. 8, pp. 2458-2462.

Jeffreys, A.J. 1985, "Individual-specific 'fingerprints' of human DNA", *Nature,* vol. 316, pp. 4.

Jiang, Y., Wang, L., Li, Z., Zhu, J., Peng, D., Su, Q., Mao, J., Wang, H., Liang, W. & Zhang, L. 2015, "A 21-plex DIP panel's application in multinational Chinese population", *Forensic Science International: Genetics Supplement Series,* vol. 5, pp. e537-e538.

Jobling, M., Hollox, E., Hurles, M., Kivisild, T. & Tyler-Smith, C. 2014, "Chapter3: Human Genome Variation" in *Human Evolutionary Genetics*, ed. Garland Science, Taylor & Francis Group, LLC, Second Edition edn, pp. 54.

Jobling, M.A. & Gill, P. 2004, "Encoded evidence: DNA in forensic analysis", *Nature Reviews Genetics,* vol. 5, no. 10, pp. 739-751.

Johnson, D.W., Pieniazek, N.J., Griffin, D.W., Misener, L. & Rose, J.B. 1995, "Development of a PCR protocol for sensitive detection of Cryptosporidium oocysts in water samples", *Applied and Environmental Microbiology,* vol. 61, no. 11, pp. 3849-3855.

Kaiser, C., Bachmeier, B., Conrad, C., Nerlich, A., Bratzke, H., Eisenmenger, W. & Peschel, O. 2008, "Molecular study of time dependent changes in DNA stability in soil buried skeletal residues", *Forensic science international,* vol. 177, no. 1, pp. 32-36.

Kalmar, T., Bachrati, C.Z., Marcsik, A. & Rasko, I. 2000, "A simple and efficient method for PCR amplifiable DNA extraction from ancient bones", *Nucleic acids research,* vol. 28, no. 12, pp. E67.

Kamarudin, K.H. & Ngah, I. 2007, *Pembangunan Mapan Orang Asli (Sustainable Development of the Orang Asli),* Penerbit UTM Press.

Karlsson, A.O., Holmlund, G., Egeland, T. & Mostad, P. 2007, "DNA-testing for immigration cases: the risk of erroneous conclusions", *Forensic science international,* vol. 172, no. 2, pp. 144-149.

Kasu, M. & Shires, K. 2015, "The validation of forensic DNA extraction systems to utilize soil contaminated biological evidence", *Legal medicine,* vol. 17, no. 4, pp. 232-238.

Katcher, H.L. & Schwartz, I. 1994, "A distinctive property of Tth DNA polymerase: enzymatic amplification in the presence of phenol", *BioTechniques,* vol. 16, no. 1, pp. 84-92.

Kee, B., Chua, K., Lee, P. & Lian, L. 2012, "Population data of six Alu insertions in indigenous groups from Sabah, Malaysia", *Annals of Human Biology,* vol. 39, no. 6, pp. 505-510.

Kelly, H., Bright, J.A., Curran, J.M. & Buckleton, J. 2012, "Modelling heterozygote balance in forensic DNA profiles", *Forensic Science International: Genetics,* vol. 6, no. 6, pp. 729-734.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. & Haussler, D. 2002, "The human genome browser at UCSC", *Genome research,* vol. 12, no. 6, pp. 996-1006.

Kermekchiev, M.B., Kirilova, L.I., Vail, E.E. & Barnes, W.M. 2009, "Mutants of Taq DNA polymerase resistant to PCR inhibitors allow DNA amplification from whole blood and crude soil samples", *Nucleic acids research,* vol. 37, no. 5, pp. e40.

Khan, G., Kangro, H.O., Coates, P.J. & Heath, R.B. 1991, "Inhibitory effects of urine on the polymerase chain reaction for cytomegalovirus DNA", *Journal of clinical pathology,* vol. 44, no. 5, pp. 360-365.

Khorasanizadeh, H., Parkkinen, J., Parthiban, R. & Moore, J.D. 2015, "Energy and economic benefits of LED adoption in Malaysia", *Renewable and Sustainable Energy Reviews,* vol. 49, pp. 629-637.

Kim, E.H., Lee, H.Y., Yang, I.S., Yang, W.I. & Shin, K.J. 2014, "Population data for 30 insertion–deletion markers in a Korean population", *International journal of legal medicine,* vol. 128, no. 1, pp. 51-52.

Kim, S., Labbe, R.G. & Ryu, S. 2000, "Inhibitory effects of collagen on the PCR for detection of Clostridium perfringens", *Applied and Environmental Microbiology,* vol. 66, no. 3, pp. 1213-1215.

Kis, Z., Zalán, A., Völgyi, A., Kozma, Z., Domján, L. & Pamjav, H. 2012, "Genome deletion and insertion polymorphisms (DIPs) in the Hungarian population", *Forensic Science International: Genetics,* vol. 6, no. 5, pp. e125-e126.

Kitayama, T., Fujii, K., Nakahara, H., Mizuno, N., Kasai, K., Yonezawa, N. & Sekiguchi, K. 2013, "Estimation of the detection rate in STR analysis by determining the DNA degradation ratio using quantitative PCR", *Legal medicine,* vol. 15, no. 1, pp. 1-6.

Kitts, A., Phan, L., Ward, M. & Holmes, J.B. 2014, "The database of short genetic variation (dbSNP)", .

Klein, D. 2002, "Quantification using real-time PCR technology: applications and limitations", *Trends in molecular medicine,* vol. 8, no. 6, pp. 257-260.

Klein, R., Neumann, C. & Roy, R. 2015, "Detection of insertion/deletion polymorphisms from challenged samples using the Investigator DIPplex® Kit", *Forensic Science International: Genetics,* vol. 16, pp. 29-37.

Kobayashi, Y., Kobayashi, Y., Watanabe, T., Shaff, J.E., Ohta, H., Kochian, L.V., Wagatsuma, T., Kinraide, T.B. & Koyama, H. 2013, "Molecular and physiological analysis of Al(3)(+) and H(+) rhizotoxicities at moderately acidic conditions", *Plant Physiology,* vol. 163, no. 1, pp. 180-192.

Köchl, S., Niederstätter, H. & Parson, W. 2005, "DNA extraction and quantitation of forensic samples using the phenol-chloroform method and real-time PCR.", *Methods in molecular biology (Clifton, N.J.),* vol. 297, pp. 13-30.

Kontanis, E.J. & Reed, F.A. 2006, "Evaluation of Real-Time PCR Amplification Efficiencies to Detect PCR Inhibitors", *Journal of forensic sciences,* vol. 51, no. 4, pp. 795-804.

Koons, B.W., Sobieralski, C.A., Comey, C.T., Baechtel, E.S., Smerick, J.B., Presley, K.W. & Stanley, D.M. 1994, "DNA extraction strategies for amplified fragment length polymorphism analysis", *journal of Forensic Science,* vol. 39, no. 5, pp. 1254-1269.

Kovatsi, L., Nikou, D., Triantaphyllou, S., Njau, S.N., Voutsaki, S. & Kouidou, S. 2009, "DNA repair enables sex identification in genetic material from human teeth", *Hippokratia,* vol. 13, no. 3, pp. 165-168.

Krause, J., Fu, Q., Good, J.M., Viola, B., Shunkov, M.V., Derevianko, A.P. & Pääbo, S. 2010, "The complete mitochondrial DNA genome of an unknown hominin from southern Siberia", *Nature,* vol. 464, no. 7290, pp. 894-897.

Kreader, C.A. 1996, "Relief of amplification inhibition in PCR with bovine serum albumin or T4 gene 32 protein", *Applied and Environmental Microbiology,* vol. 62, no. 3, pp. 1102-1106.

Krenke, B.E., Tereba, A., Anderson, S.J., Buel, E., Culhane, S., Finis, C.J., Tomsey, C.S., Zachetti, J.M., Masibay, A. & Rabbach, D.R. 2002, "Validation of a 16-locus fluorescent multiplex system", *Journal of forensic sciences,* vol. 47, no. 4, pp. 773-785.

Krishnamurthy, V., Manoj, R. & Pagare, S.S. 2011, "Understanding the Basic of DNA Fingerprinting in Forensic Science", *Journal of Indian Academy of Oral Medicine and Radiology,* vol. 23(4), pp. 613-616.

Lantz, P.G., Matsson, M., Wadström, T. & 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,* vol. 28, no. 3, pp. 159-167.

Larue, B.L., Ge, J., King, J.L. & Budowle, B. 2012, "A validation study of the Qiagen Investigator DIPplex (R) kit; an INDEL-based assay for human identification", *International journal of legal medicine,* vol. 126, no. 4, pp. 533-540.

Leat, N., Ehrenrieich, L., Benjeddou, M. & Davison, S. 2004, "Developments in the use of Y-chromosome markers in forensic genetics", *African Journal of Biotechnology,* vol. 3, no. 12, pp. 637-642.

Leclair, B., Sgueglia, J.B., Wojtowicz, P.C., Juston, A.C., Fregeau, C.J. & Fourney, R.M. 2003, "STR DNA typing: increased sensitivity and efficient sample consumption using reduced PCR reaction volumes", *Journal of forensic sciences,* vol. 48, no. 5, pp. 1001-1013.

Lee, A.B. & Cooper, T.A. 1995, "Improved direct PCR screen for bacterial colonies: wooden toothpicks inhibit PCR amplification", *BioTechniques,* vol. 18, no. 2, pp. 225-226.

Li, C., Zhao, S., Zhang, S., Li, L., Liu, Y., Chen, J. & Xue, J. 2011, "Genetic polymorphism of 29 highly informative InDel markers for forensic use in the Chinese Han population", *Forensic Science International-Genetics,* vol. 5, no. 1, pp. E27-E30.

Lienert, K. & Fowler, J.C. 1992, "Analysis of mixed human/microbial DNA samples: a validation study of two PCR AMP-FLP typing methods", *BioTechniques,* vol. 13, no. 2, pp. 276-281.

Lim, K.B., Jeevan, N.H., Jaya, P., Othman, M.I. & Lee, Y.H. 2001, "STR data for the AmpFISTR Profiler loci from the three main ethnic population groups (Malay, Chinese, Indian) in Malaysia.", *Forensic science international,* vol. 119, pp. 109-112.

Listman, J.B., Malison, R.T., Sughondhabirom, A., Yang, B.Z., Raaum, R.L., Thavichachart, N., Sanichwankul, K., Kranzler, H.R., Tangwonchai, S., Mutirangura, A., Disotell, T.R. & Gelernter, J. 2007, "Demographic changes and marker properties affect detection of human population differentiation", *BMC genetics,* vol. 8, pp. 21.

Magalhães, M.C.A., Amorim, A., Gusmão, L., Gorostiza, A., González-Martín, A. & Pereira, R. 2015, "Autosomal indels distribution in Metropolitan Manila, Philippines", *Forensic Science International: Genetics Supplement Series,* vol. 5, pp. e451-e453.

Mahony, J., Chong, S., Jang, D., Luinstra, K., Faught, M., Dalby, D., Sellors, J. & Chernesky, M. 1998, "Urine specimens from pregnant and nonpregnant 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,* vol. 36, no. 11, pp. 3122-3126.

Manabe, S., Mori, Y., Kawai, C., Ozeki, M. & Tamaki, K. 2013, "Mixture interpretation: Experimental and simulated reevaluation of qualitative analysis", *Legal medicine (Tokyo, Japan),* vol. 15, no. 2, pp. 66-71.

Manta, F., Caiafa, A., Pereira, R., Silva, D., Amorim, A., Carvalho, E.F. & Gusmao, 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,* vol. 6, no. 5, pp. 658-661.

Marasco, E., Ross, A., Dawson, J., Moroose, T. & Ambrose, T. 2012, "Detecting STR peaks in degraded DNA samples", *Proceedings of the 4th International Conference of Bioinformatics and Computational Biology*, pp. 1.

Markoulatos, P., Siafakas, N. & Moncany, M. 2002, "Multiplex polymerase chain reaction: a practical approach", *Journal of clinical laboratory analysis,* vol. 16, no. 1, pp. 47-51.

Marshall, P.L., King, J.L., Lawrence, N.P., Lazarev, A., Gross, V.S. & Budowle, B. 2013, "Pressure cycling technology (PCT) reduces effects of inhibitors of the PCR", *International journal of legal medicine,* vol. 127, no. 2, pp. 321-333.

Martín, P., de Simón, L.F., Luque, G., Farfán, M.J. & Alonso, A. 2014, "Improving DNA data exchange: validation studies on a single 6 dye STR kit with 24 loci", *Forensic Science International: Genetics,* vol. 13, pp. 68-78.

Martin, P., Garcia, O., Heinrichs, B., Yurrebaso, I., Aguirre, A. & Alonso, A. 2013, "Population genetic data of 30 autosomal indels in Central Spain and the Basque Country populations", *Forensic Science International-Genetics,* vol. 7, no. 2, pp. E27-E30.

Martínez-Cortés, G., García-Aceves, M., Favela-Mendoza, A., Muñoz-Valle, J., Velarde-Felix, J. & Rangel-Villalobos, H. 2015, "Forensic parameters of the Investigator DIPplex kit (Qiagen) in six Mexican populations", *International journal of legal medicine,* , pp. 1-3.

Maruyama, S., Minaguchi, K., Takezaki, N. & Nambiar, P. 2008, "Population data on 15 STR loci using AmpFISTR Identifiler kit in a Malay population living in and around Kuala Lumpur, Malaysia", *Legal Medicine,* vol. 10, pp. 160-162.

McCord, B., Opel, K., Chung, D., Drabek, J., Tatarek, N., Jantz, L.M. & Butler, J. 2005, "The development of miniplex primer sets for the analysis of degraded DNA", *Defense and Security*International Society for Optics and Photonics, , pp. 617.

McLaren, R. 2007, "PowerPlex® 16 versus Identifiler® Systems: Sensitivity and effects of inhibitors", *Application Notes Promega Corporation,* [Online], vol. 156, . Available from: https://www.promega.co.uk/~/media/files/resources/application%20notes/genetic%20identity/an156%20powerplex%2016%20versus%20identifiler%20systems%20sensitivity%20and%20effects%20of%20inhibitors.pdf?la=en.

Meng, H.T., Zhang, Y.D., Shen, C.M., Yuan, G.L., Yang, C.H., Jin, R., Yan, J.W., Wang, H.D., Liu, W.J. & Jing, H. 2015, "Genetic polymorphism analyses of 30 InDels in Chinese Xibe ethnic group and its population genetic differentiations with other groups", *Scientific reports,* vol. 5.

Mills, R.E., Luttig, C.T., Larkins, C.E., Beauchamp, A., Tsui, C., Pittard, W.S. & Devine, S.E. 2006, "An initial map of insertion and deletion (INDEL) variation in the human genome", *Genome research,* vol. 16, no. 9, pp. 1182-1190.

Monteiro, L., Bonnemaison, D., Vekris, A., Petry, K.G., Bonnet, J., Vidal, R., Cabrita, J. & Megraud, F. 1997, "Complex polysaccharides as PCR inhibitors in feces: Helicobacter pylori model", *Journal of clinical microbiology,* vol. 35, no. 4, pp. 995-998.

Moura-Neto, R., Silva, R., Mello, I., Nogueira, T., Al-Deib, A., LaRue, B., King, J. & Budowle, B. 2015, "Evaluation of a 49 InDel Marker HID panel in two specific populations of South America and one population of Northern Africa", *International journal of legal medicine,* vol. 129, no. 2, pp. 245-249.

Mullaney, J.M., Mills, R.E., Pittard, W.S. & Devine, S.E. 2010, "Small insertions and deletions (INDELs) in human genomes", *Human molecular genetics,* vol. 19, no. R2, pp. R131-6.

Mullikin, 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. and Ainscough, R.M.R., 2000. An SNP map of human chromosome 22. *Nature*, *407*(6803), pp.516-520.

Murthy, V., Jia, L.F., Samuel, V.P. & Kademane, K. 2015, "Forensic identification by using insertion deletion Polymorphisms", *International Journal of Human Genetics,* vol. 15, no. 2, pp. 55-59.

Nachman, M.W. & Crowell, S.L. 2000, "Estimate of the mutation rate per nucleotide in humans", *Genetics,* vol. 156, no. 1, pp. 297-304.

NanoDrop 2000/2000c Spectrophotometer V1.0 User Manual 2009, *Thermo Scientific*, Thermo Fisher Scientific, USA.

Nazir, M.S. 2012, *DNA Persistence and Preservation Following Environmental Insult*, University of Central Lancashire.

Nazir, M.S., Iyavoo, S., Alimat, S., Zahra, N., Sanqoor, S.H., Smith, J.A., Moffatt, C. & Goodwin, W. 2013, "Development of a multiplex system to assess DNA persistence in taphonomic studies", *Electrophoresis,* vol. 34, no. 24, pp. 3352-3360.

Neuvonen, A.M., Palo, J.U., Hedman, M. & Sajantila, A. 2012, "Discrimination power of Investigator DIPplex loci in Finnish and Somali populations", *Forensic Science International: Genetics,* vol. 6, no. 4, pp. e99-e102.

Nomura, T., Honda, T. & Mukai, F. 2001, "Inbreeding and effective population size of Japanese Black cattle.", *Journal of animal science,* vol. 79, no. 2, pp. 366-370.

Norhalifah, H.K., Syaza, F.H., Chambers, G.K. & Edinur, H.A. 2016, "The genetic history of Peninsular Malaysia", *Gene,* vol. 586, no. 1, pp. 129-135.

Normile, D. 2009, "Genetics. SNP study supports southern migration route to Asia", *Science (New York, N.Y.),* vol. 326, no. 5959, pp. 1470.

Nunotani, M., Shiozaki, T., Sato, N., Kamei, S., Takatsu, K., Hayashi, T., Ota, M. & Asamura, H. 2015, "Analysis of 30 insertion–deletion polymorphisms in the Japanese population using the Investigator DIPplex® kit", *Legal medicine,* vol. 17, no. 6, pp. 467-470.

O'Connor, K.L., Hill, C.R., Vallone, P.M. & Butler, J.M. 2011, "Linkage disequilibrium analysis of D12S391 and vWA in US population and paternity samples", *Forensic Science International: Genetics,* vol. 5, no. 5, pp. 538-540.

Oka, K., Asari, M., Omura, T., Yoshida, M., Maseda, C., Yajima, D., Matsubara, K., Shiono, H.i., Matsuda, M. & Shimizu, K. 2014, "Genotyping of 38 insertion/deletion polymorphisms for human identification using universal fluorescent PCR", *Molecular and cellular probes,* vol. 28, no. 1, pp. 13-18.

Opel, K.L., Chung, D. & McCord, B.R. 2010, "A Study of PCR Inhibition Mechanisms Using Real Time PCR*,†", *Journal of forensic sciences,* vol. 55, no. 1, pp. 25-33.

Opel, K.L., Chung, D.T., Drábek, J., Butler, J.M. & McCord, B.R. 2007, "Developmental Validation of Reduced-Size STR Miniplex Primer Sets", *Journal of forensic sciences,* vol. 52, no. 6, pp. 1263-1271.

Ostell, J. 2014, "The Entrez search and retrieval system", .

Othman, M.I., Seah, L.H., Panneerchelvam, S. & Nor, N.M. 2004, "Allele Frequencies for the PowerPlex® 16 STR Loci in Javanese Population in Malaysia", *Journal of forensic sciences,* vol. 49, no. 1, pp. 190-191.

Ozeki, M., Kato, H. & Tamaki, K. 2015, "Development of a novel miniplex DNA identification system for the Japanese population", *Legal medicine,* vol. 17, no. 3, pp. 205-209.

Pajnič, I.Z., Debska, M., Pogorelc, B.G., Mohorčič, K.V., Balažic, J., Zupanc, T., Štefanič, B. & Geršak, K. 2016, "Highly efficient automated extraction of DNA from old and contemporary skeletal remains", *Journal of forensic and legal medicine,* vol. 37, pp. 78-86.

Palha, T.J.B.F., Rodrigues, E.M.R., Cavalcante, G.C., Marrero, A., de Souza, I.R., Uehara, C.J.S., da Motta, C.H.A.S., Koshikene, D., da Silva, D.A. & de Carvalho, E.F. 2015, "Population genetic analysis of insertion–deletion polymorphisms in a Brazilian population using the Investigator DIPplex kit", *Forensic Science International: Genetics,* vol. 19, pp. 10-14.

Pepinski, W., Abreu-Glowacka, M., Koralewska-Kordel, M., Michalak, E., Kordel, K., Niemcunowicz-Janica, A., Szeremeta, M. & Konarzewska, M. 2013, "Population genetics of 30 INDELs in populations of Poland and Taiwan", *Molecular biology reports,* vol. 40, no. 7, pp. 4333-4338.

Pereira, R., Phillips, C., Pinto, N., Santos, C., dos Santos, S. E. B., Amorim, A., Carracedo, Á. & Gusmão, L. 2012, "Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing", *PLoS One,* vol. 7, no. 1, pp. e29684.

Pereira, V., Moncada, E., Diez, I.E., Tomas, C., Amorim, A., Morling, N., Gusmão, L. & Prata, M.J. 2011, "Genetic characterization of Somali and Iraqi populations using a set of 33 X-chromosome Indels", *Forensic Science International: Genetics Supplement Series,* vol. 3, no. 1, pp. e137-e138.

Pereira, R. & Gusmao, L. 2012, "Capillary Electrophoresis of 38 Noncoding Biallelic Mini-Indels for Degraded Samples and as Complementary Tool in Paternity Testing", *Dna Electrophoresis Protocols for Forensic Genetics,* vol. 830, pp. 141-157.

Pereira, R., Pereira, V., Gomes, I., Tomas, C., Morling, N., Amorim, A., Prata, M.J., Carracedo, A. & Gusmao, L. 2012, "A method for the analysis of 32 X chromosome insertion deletion polymorphisms in a single PCR", *International journal of legal medicine,* vol. 126, no. 1, pp. 97-105.

Pereira, R., Phillips, C., Alves, C., Amorim, A., Carracedo, A. & Gusmao, L. 2009, "A new multiplex for human identification using insertion/deletion polymorphisms", *Electrophoresis,* vol. 30, no. 21, pp. 3682-3690.

Petricevic, S., Whitaker, J., Buckleton, J., Vintiner, S., Patel, J., Simon, P., Ferraby, H., Hermiz, W. & Russell, A. 2010, "Validation and development of interpretation guidelines for low copy number (LCN) DNA profiling in New Zealand using the AmpFlSTR® SGM Plus™ multiplex", *Forensic Science International: Genetics,* vol. 4, no. 5, pp. 305-310.

Phillips, C., Ballard, D., Gill, P., Court, D.S., Carracedo, A. & Lareu, M. 2012, "The recombination landscape around forensic STRs: accurate measurement of genetic distances between syntenic STR pairs using HapMap high density SNP data", *Forensic Science International: Genetics,* vol. 6, no. 3, pp. 354-365.

Phillips, C., Fondevila, M., García-Magariños, M., Rodriguez, A., Salas, A., Carracedo, A. & Lareu, M.V. 2008, "Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers", *Forensic Science International: Genetics,* vol. 2, no. 3, pp. 198-204.

Phillips, C., Salas, A., Sánchez, J.J., Fondevila, M., Gómez-Tato, A., Álvarez-Dios, J., Calaza, M., de Cal, M.C., Ballard, D., Lareu, M.V. & Carracedo, A. 2007, "Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs", *Forensic Science International: Genetics,* vol. 1, no. 3-4, pp. 273-280.

Pimenta, J. & Pena, S. 2010, "Efficient human paternity testing with a panel of 40 short insertion–deletion polymorphisms", *Genet Mol Res,* vol. 9, no. 1, pp. 601-607.

Pinto, N., Magalhães, M., Conde-Sousa, E., Gomes, C., Pereira, R., Alves, C., Gusmão, L. & Amorim, A. 2013, "Assessing paternities with inconclusive STR results: the suitability of bi-allelic markers", *Forensic Science International: Genetics,* vol. 7, no. 1, pp. 16-21.

Pontes, M.L., Fondevila, M., Laréu, M.V. & Medeiros, R. 2015, "SNP Markers as Additional Information to Resolve Complex Kinship Cases", *Transfusion Medicine and Hemotherapy,* vol. 42, no. 6, pp. 385-388.

Poulsen, L., Farzad, M.S., Børsting, C., Tomas, C., Pereira, V. & Morling, N. 2015, "Population and forensic data for three sets of forensic genetic markers in four ethnic groups from Iran: Persians, Lurs, Kurds and Azeris", *Forensic Science International: Genetics,* vol. 17, pp. 43-46.

Powell, H., Gooding, C., Garrett, S., Lund, B. & McKee, R. 1994, "Proteinase inhibition of the detection of Listeria monocytogenes in milk using the polymerase chain reaction", *Letters in applied microbiology,* vol. 18, no. 1, pp. 59-61.

PowerPlex® 16 System Technical Manual 2016, , Promega Corporation, USA.

QIAamp® DNA Mini and Blood Mini Handbook April 2010, *Qiagen®,* Third Edition edn, © 1999–2010 QIAGEN.

Quantifiler® Kits User Manual 2012, *Quantifiler® Human DNA Quantification Kit and Quantifiler® Y Human Male DNA Quantification Kit- Applied Biosystems*, Life Technologies Corporation, United States.

Qubit® dsDNA HS Assay Kits Manual 2015, *Life Technologies*, Thermo Fisher Scientific.

Rådström, P., Knutsson, R., Wolffs, P., Lövenklev, M. & Löfström, C. 2004, "Pre-PCR processing", *Molecular biotechnology,* vol. 26, no. 2, pp. 133-146.

Rajeevan, H., Soundararajan, U., Pakstis, A.J. & Kidd, K.K. 2012, "Introducing the Forensic Research/Reference on Genetics knowledge base, FROG-kb", *Investigative genetics,* vol. 3, no. 1, pp. 1.

Raymond, M. & Rousset, F. 1995, "An exact test for population differentiation", *Evolution,* vol. 49, no. 6, pp. 1280-1283.

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. & 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,* vol. 6, no. 4, pp. 469-476.

Rosenbloom, K.R., Armstrong, J., Barber, G.P., Casper, J., Clawson, H., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M., Harte, R.A., Heitner, S., Hickey, G., Hinrichs, A.S., Hubley, R., Karolchik, D., Learned, K., Lee, B.T., Li, C.H., Miga, K.H., Nguyen, N., Paten, B., Raney, B.J., Smit, A.F., Speir, M.L., Zweig, A.S., Haussler, D., Kuhn, R.M. & Kent, W.J. 2015, "The UCSC Genome Browser database: 2015 update", *Nucleic acids research,* vol. 43, no. Database issue, pp. D670-81.

Rossen, L., Nørskov, P., Holmstrøm, K. & Rasmussen, O.F. 1992, "Inhibition of PCR by components of food samples, microbial diagnostic assays and DNA-extraction solutions", *International journal of food microbiology,* vol. 17, no. 1, pp. 37-45.

Rouhibakhsh, A., Priya, J., Periasamy, M., Haq, Q. & Malathi, V. 2008, "An improved DNA isolation method and PCR protocol for efficient detection of multicomponents of begomovirus in legumes", *Journal of virological methods,* vol. 147, no. 1, pp. 37-42.

Roy, S., Rahman, I.A., Santos, J.H. & Ahmed, M.U. 2016, "Meat species identification using DNA-redox electrostatic interactions and non-specific adsorption on graphene biochips", *Food Control,* vol. 61, pp. 70-78.

Rychlik, W. 1993, "Selection of primers for polymerase chain reaction", *PCR Protocols: Current Methods and Applications,* , pp. 31-40.

Saat, I. 2009, "Sejarah Panjang Hubungan Malaysia-China.", .

Saiz, M., Andre, F., Pisano, N., Sandberg, N., Bertoni, B. & Pagano, S. 2014, "Allelic frequencies and statistical data from 30 INDEL loci in Uruguayan population", *Forensic science international.Genetics,* vol. 9, pp. e27-9.

Sanchez, J.J., Phillips, C., Borsting, C., Balogh, K., Bogus, M., Fondevila, M., Harrison, C.D., Musgrave-Brown, E., Salas, A., Syndercombe-Court, D., Schneider, P.M., Carracedo, A. & Morling, N. 2006, "A multiplex assay with 52 single nucleotide polymorphisms for human identification", *Electrophoresis,* vol. 27, no. 9, pp. 1713-1724.

Sanga, M., Boiso, L., Lindsten, H., Rådström, P., Ansell, R. & Hedman, J. 2015, "A panel of PCR-inhibitory reference materials for quality evaluation of multiplex STR analysis kits", *Forensic Science International: Genetics Supplement Series,* vol. 5, pp. e317-e319.

Santos, N.P.C., Ribeiro-Rodrigues, E.M., Ribeiro-dos-Santos, Â.K.C., Pereira, R., Gusmão, L., Amorim, A., Guerreiro, J.F., Zago, M.A., Matte, C., & Mara, H. 2010, "Assessing individual interethnic admixture and population substructure using a 48–insertion-deletion (INSEL) ancestry-informative marker (AIM) panel", *Human mutation,* vol. 31, no. 2, pp. 184-190.

Santos, V.R., Pena, H.B. & Pena, S.D. 2015, "A multiplex panel of short-amplicon insertion-deletion DNA polymorphisms for forensic analysis", *Genetics and molecular research : GMR,* vol. 14, no. 2, pp. 2947-2952.

Satsangi, J., Jewell, D., Welsh, K., Bunce, M. & Bell, J. 1994, "Effect of heparin on polymerase chain reaction", *The Lancet,* vol. 343, no. 8911, pp. 1509-1510.

Saulnier, P. & Andremont, A. 1992, "Detection of genes in feces by booster polymerase chain reaction", *Journal of clinical microbiology,* vol. 30, no. 8, pp. 2080-2083.

Sawaya, S., Bagshaw, A., Buschiazzo, E., Kumar, P., Chowdhury, S., Black, M.A. & Gemmell, N. 2013, "Microsatellite Tandem Repeats Are Abundant in Human Promoters and Are Associated with Regulatory Elements", *PLoS ONE,* vol. 8, no. 2.

Scheible, M., Loreille, O., Just, R. & Irwin, J. 2014, "Short tandem repeat typing on the 454 platform: strategies and considerations for targeted sequencing of common forensic markers", *Forensic Science International: Genetics,* vol. 12, pp. 107-119.

Schiffner, L.A., Bajda, E.J., Prinz, M., Sebestyen, J., Shaler, R. & Caragine, T.A. 2005, "Optimization of a simple, automatable extraction method to recover sufficient DNA from low copy number DNA samples for generation of short tandem repeat profiles", *Croatian medical journal,* vol. 46, no. 4, pp. 578.

Schneider, P.M. 2012, "Beyond STRs: The Role of Diallelic Markers in Forensic Genetics", *Transfusion Medicine and Hemotherapy,* vol. 39, no. 3, pp. 176-180.

Scholz, M., Giddings, I. & Pusch, C.M. 1998, "A polymerase chain reaction inhibitor of ancient hard and soft tissue DNA extracts is determined as human collagen type I", *Analytical Biochemistry,* vol. 259, no. 2, pp. 283-286.

Schoske, R., Vallone, P.M., Ruitberg, C.M. & Butler, J.M. 2003, "Multiplex PCR design strategy used for the simultaneous amplification of 10 Y chromosome short tandem repeat (STR) loci", *Analytical and Bioanalytical Chemistry,* vol. 375, no. 3, pp. 333-343.

Schrader, C., Schielke, A., Ellerbroek, L. & Johne, R. 2012, "PCR inhibitors–occurrence, properties and removal", *Journal of applied microbiology,* vol. 113, no. 5, pp. 1014-1026.

Scientific Working Group on DNA Analysis Methods (SWGDAM), United States of America & Journal: 2004, "Revised Validation Guidelines", .

Seah, L.H., Jeevan, N.H., Othman, M.I., Jaya, P., Ooi, Y.S., Wong, P.C. & Kee, S.S. 2003, "STR data for the AmpFISTR Identifiler loci in three ethnic groups (Malay, Chinese, Indian) of the Malaysian population", *Forensic science international,* vol. 138, pp. 134-137.

Seah, L.H., Phoon, Y.K., Wee, B.H. & Asam, A. 2016, "Sensitivity study of 30 Indels in the Investigator DIPplex Kit for forensic casework applications.", *Forensic Asia,* [Online], no. 7, pp. 18/9/2016-20-21. Available from: https://www.researchgate.net/publication/303652315_Sensitivity_Study_of_30_Indels_in_the_InvestigatorR_DIPplex_Kit_for_Forensic_Casework_Applications.

Seo, S.B., Lee, H.Y., Zhang, A.H., Kim, H.Y., Shin, D.H. & Lee, S.D. 2012, "Effects of humic acid on DNA quantification with Quantifiler® Human DNA Quantification kit and short tandem repeat amplification efficiency", *International journal of legal medicine,* vol. 126, no. 6, pp. 961-968.

Seong, K.M., Park, J.H., Hyun, Y.S., Kang, P.W., Choi, D.H., Han, M.S., Park, K.W. & Chung, K.W. 2014, "Population genetics of insertion–deletion polymorphisms in South Koreans using Investigator DIPplex kit", *Forensic Science International: Genetics,* vol. 8, no. 1, pp. 80-83.

Shafique, M., Shahzad, M.S., Rahman, Z., Shan, M.A., Perveen, R., Shahzad, M., Hussain, M., Shahid, A.A. & Husnain, T. 2016, "Development of new PCR multiplex system by the simultaneous detection of 10 miniSTRs, SE33, Penta E, Penta D, and four Y-STRs", *International journal of legal medicine,* , pp. 1-11.

Sharizah A. 2013,
Application of single nucleotide polymorphisms (SNPs) to forensic casework in Malaysia, University of Central Lancashire.

Sherry, S.T., Ward, M. & Sirotkin, K. 1999, "dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation", *Genome research,* vol. 9, no. 8, pp. 677-679.

Shi, M., Liu, Y., Bai, R., Jiang, L., Lv, X. & Ma, S. 2015, "Population data of 30 insertion–deletion markers in four Chinese populations", *International journal of legal medicine,* vol. 129, no. 1, pp. 53-56.

Simon, M.C., Gray, D.I. & Cook, N. 1996, "DNA Extraction and PCR Methods for the Detection of Listeria monocytogenes in Cold-Smoked Salmon", *Applied and Environmental Microbiology,* vol. 62, no. 3, pp. 822-824.

Simonson, T.S., Xing, J., Barrett, R., Jerah, E., Loa, P., Zhang, Y., Watkins, W.S., Witherspoon, D.J., Huff, C.D. & Woodward, S. 2011, "Ancestry of the Iban is predominantly Southeast Asian: genetic evidence from autosomal, mitochondrial, and Y chromosomes", *PLoS One,* vol. 6, no. 1, pp. 1-8.

Singh, N.P. 2016, "The comet assay: Reflections on its development, evolution and applications", *Mutation Research/Reviews in Mutation Research,* vol. 767, pp. 23-30.

Slatkin, M. 1995, "A measure of population subdivision based on microsatellite allele frequencies", *Genetics,* vol. 139, no. 1, pp. 457-462.

Smith, P.J. & Ballantyne, J. 2007, "Simplified Low-Copy-Number DNA Analysis by Post-PCR Purification", *Journal of forensic sciences,* vol. 52, no. 4, pp. 820-829.

Suadi, Z., Siew, L.C., Tie, R., Hui, W.B., Asam, A., Thiew, S.H. & Boon, L.K. 2007, " STR Data for the AmpFISTR Identifiler Loci from the three main ethnic indigenous population groups (Iban, Bidayuh and Melanau) in Sarawak ", *Journal of forensic sciences,* vol. 52, no. 1, pp. 231-234.

Sutlovic, D., Gamulin, S., Definis, G.M., Gugic, D. & Andjelinovic, S. 2008, "Interaction of humic acids with human DNA: proposed mechanisms and kinetics", *Electrophoresis,* vol. 29, no. 7, pp. 1467-1472.

Swango, K.L., Timken, M.D., Chong, M.D. & Buoncristiani, M.R. 2006, "A quantitative PCR assay for the assessment of DNA degradation in forensic samples", *Forensic science international,* vol. 158, no. 1, pp. 14-26.

Tamariz, J., Voynarovska, K., Prinz, M. & Caragine, T. 2006, "The application of ultraviolet irradiation to exogenous sources of DNA in plasticware and water for the amplification of low copy number DNA", *Journal of forensic sciences,* vol. 51, no. 4, pp. 790-794.

Tamura, T., Osawa, M., Ochiai, E., Suzuki, T. & Nakamura, T.i. 2015, "Evaluation of advanced multiplex short tandem repeat systems in pairwise kinship analysis", *Legal medicine,* vol. 17, no. 5, pp. 320-325.

Tan, S. 2001, *Genetic relationships among sixteen ethnic groups from Malaysia and Southeast Asia,* Singapore: World Scientific.

Taylor, D., Bright, J.A., McGoven, C., Hefford, C., Kalafut, T. & Buckleton, J. 2016, "Validating multiplexes for use in conjunction with modern interpretation strategies", *Forensic Science International: Genetics,* vol. 20, pp. 6-19.

Tereba, A. 1999, "Tools for analysis of population statistics", *Profiles in DNA,* vol. 2, no. 3, pp. 14-16.

Tereba, A., Masibay, A., Krenke, B.E., Finis, C.J., Tomsey, C.S., Sprecher, C.J., Rabbach, D.R., Amiott, E.A., Buel, E. & Zachetti, J.M. 2002, "Validation of a 16-locus fluorescent multiplex system", *Journal of Forensic Science,* vol. 47, no. 4, pp. 1-13.

Thompson, R.E., Duncan, G. & McCord, B.R. 2014, "An Investigation of PCR Inhibition Using Plexor®-Based Quantitative PCR and Short Tandem Repeat Amplification", *Journal of forensic sciences,* vol. 59, no. 6, pp. 1517-1529.

Thornton, C.G. & Passen, S. 2004, "Inhibition of PCR amplification by phytic acid, and treatment of bovine fecal specimens with phytase to reduce inhibition", *Journal of microbiological methods,* vol. 59, no. 1, pp. 43-52.

Tilmar, A. 2010, *Populations and statistics in Forensic Genetics*, Linkoping University, Sweden.

Tomas, C., Poulsen, L., Drobnic, K., Ivanova, V., Jankauskiene, J., Bunokiene, D., Børsting, C. & Morling, N. 2016, "The autosomal insertion-deletion polymorphisms analyzed using the Investigator® DIPplex Kit in populations from Iraq, Lithuania, Slovenia, and Turkey", *Forensic Science International: Genetics,* vol. 25, pp. 142-144.

Tomas, C., Skitsa, I., Steinmeier, E., Poulsen, L., Ampati, A., Børsting, C. & Morling, N. 2015, "Results for five sets of forensic genetic markers studied in a Greek population sample", *Forensic Science International: Genetics,* vol. 16, pp. 132-137.

Torres, S.R.R., Uehara, C.J.S., Sutter-Latorre, A.F., de Almeida, B.S., Sauerbier, T.S., Muniz, Y.C.N., Marrero, A.R. & de Souza, I.R. 2014, "Population genetic data and forensic parameters of 30 autosomal InDel markers in Santa Catarina State population, Southern Brazil", *Molecular biology reports,* vol. 41, no. 8, pp. 5429-5433.

Tsai, Y.L. & Olson, B.H. 1992b, "Detection of low numbers of bacterial cells in soils and sediments by polymerase chain reaction", *Applied and Environmental Microbiology,* vol. 58, no. 2, pp. 754-757.

Tsai, Y.L. & Olson, B.H. 1992a, "Rapid method for separation of bacterial DNA from humic substances in sediments for polymerase chain reaction", *Applied and Environmental Microbiology,* vol. 58, no. 7, pp. 2292-2295.

Wadowsky, R.M., Laus, S., Libert, T., States, S.J. & Ehrlich, G.D. 1994, "Inhibition of PCR-based assay for Bordetella pertussis by using calcium alginate fiber and aluminum shaft components of a nasopharyngeal swab", *Journal of clinical microbiology,* vol. 32, no. 4, pp. 1054-1057.

Wang, D.Y., Chang, C.W., Lagacé, R.E., Calandro, L.M. & Hennessy, L.K. 2012, "Developmental validation of the AmpFℓSTR® Identifiler® Plus PCR Amplification Kit: an established multiplex assay with improved performance", *Journal of forensic sciences,* vol. 57, no. 2, pp. 453-465.

Wang, D.Y., Chang, C.W., Lagacé, R.E., Oldroyd, N.J. & Hennessy, L.K. 2011, "Development and Validation of the AmpFℓSTR® Identifiler® Direct PCR Amplification Kit: A Multiplex Assay for the Direct Amplification of Single-Source Samples", *Journal of forensic sciences,* vol. 56, no. 4, pp. 835-845.

Wang, L., Lv, M., Zaumsegel, D., Zhang, L., Liu, F., Xiang, J., Li, J., Schneider, P.M., Liang, W. & Zhang, L. 2016, "A comparative study of insertion/deletion polymorphisms applied among Southwest, South and Northwest Chinese populations using Investigator® DIPplex", *Forensic Science International: Genetics,* vol. 21, pp. 10-14.

Wang, Z., Zhang, S., Zhao, S., Hu, Z., Sun, K. & Li, C. 2014, "Population genetics of 30 insertion–deletion polymorphisms in two Chinese populations using Qiagen Investigator® DIPplex kit", *Forensic Science International: Genetics,* vol. 11, pp. e12-e14.

Watanabe, Y., Takayama, T., Hirata, K., Yamada, S., Nagai, A., Nakamura, I., Bunai, Y. & Ohya, I. 2003, "DNA typing from cigarette butts", *Legal medicine,* vol. 5, pp. S177-S179.

Watson, R. & Blackwell, B. 2000, "Purification and characterization of a common soil component which inhibits the polymerase chain reaction", *Canadian journal of microbiology,* vol. 46, no. 7, pp. 633-642.

Weber, J.L., David, D., Heil, J., Fan, Y., Zhao, C.F. & Marth, G. 2002, "Human diallelic insertion/deletion polymorphisms", *American Journal of Human Genetics,* vol. 71, no. 4, pp. 854-862.

Weir, B.S. & Cockerham, C.C. 1984, "Estimating F-statistics for the analysis of population structure.", *Evolution,* vol. 38, no. 6, pp. 1358-1370.

Weir, B.S. & Hill, W.G. 2002, "Estimating F-Statistics", *Annu.Rev.Genet,* vol. 36, pp. 721-750.

Westen, A.A., Nagel, J.H.A., Benschop, C.C.G., Weiler, N.E.C., De Jong, B.J. & Sijen, T. 2009, "Higher capillary electrophoresis injection settings as an efficient approach to increase the sensitivity of STR typing", *Journal of forensic sciences,* vol. 54, no. 3, pp. 591-598.

Westring, C.G., Kristinsson, R., Gilbert, D.M. & Danielson, P.B. 2007, "Validation of reduced-scale reactions for the Quantifiler™ human DNA kit", *Journal of forensic sciences,* vol. 52, no. 5, pp. 1035-1043.

Whitaker, J., Cotton, E. & Gill, P. 2001, "A comparison of the characteristics of profiles produced with the AMPFlSTR® SGM Plus™ multiplex system for both standard and low copy number (LCN) STR DNA analysis", *Forensic science international,* vol. 123, no. 2, pp. 215-223.

Wiegand, P. & Kleiber, M. 2001, "Less is more–length reduction of STR amplicons using redesigned primers", *International journal of legal medicine,* vol. 114, no. 4-5, pp. 285-287.

Wilson, I.G. 1997, "Inhibition and facilitation of nucleic acid amplification.", *Applied and Environmental Microbiology,* vol. 63, no. 10, pp. 3741.

Witherspoon, D.J., Wooding, S., Rogers, A.R., Marchani, E.E., Watkins, W.S., Batzer, M.A. & Jorde, L.B. 2007, "Genetic similarities within and between human populations", *Genetics,* vol. 176, no. 1, pp. 351-359.

Word, C. 2010, "What is LCN?—Definitions and Challenges", [Online], , pp. 17/09/2016. Available from: http://www.promega.co.uk/resources/profiles-in-dna/2010/what-is-lcn-definitions-and-challenges/.

Yates, A., Akanni, W., Amode, M.R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., Giron, C.G., Gordon, L., Hourlier, T., Hunt, S.E., Janacek, S.H., Johnson, N., Juettemann, T., Keenan, S., Lavidas, I., Martin, F.J., Maurel, T., McLaren, W., Murphy, D.N., Nag, R., Nuhn, M., Parker, A., Patricio, M., Pignatelli, M., Rahtz, M., Riat, H.S., Sheppard, D., Taylor, K., Thormann, A., Vullo, A., Wilder, S.P., Zadissa, A., Birney, E., Harrow, J., Muffato, M., Perry, E., Ruffier, M., Spudich, G., Trevanion, S.J., Cunningham, F., Aken, B.L., Zerbino, D.R. & Flicek, P. 2016, "Ensembl 2016", *Nucleic acids research,* vol. 44, no. D1, pp. D710-6.

Yun, L., Gao, T., Sun, K., Gu, Y. & Hou, Y. 2015, "Comparison of two online algorithm methods for forensic ancestry inference", *Forensic Science International: Genetics Supplement Series,* vol. 5, pp. e559-e560.

Zahra, N., Hadi, S. & Goodwin, W. 2012, "Development of PCR internal controls for DNA profiling with the AmpFℓSTR® SGM Plus® amplification kit", *Electrophoresis,* vol. 33, no. 18, pp. 2833-2839.

Zahra, N., Hadi, S., Smith, J.A., Iyengar, A. & Goodwin, W. 2011, "Development of internal amplification controls for DNA profiling with the AmpFISTR SGM Plus kit", *Electrophoresis,* , no. 32, pp. 1371-1378.

Zahra, N. & Goodwin, W. 2016, "The Development and Use of Internal Amplification Controls (IACs) with DNA Profiling Kits for Forensic DNA Analysis", *Methods in molecular biology (Clifton, N.J.),* vol. 1420, pp. 109-124.

Zaragoza, M.V., Fass, J., Diegoli, M., Lin, D. & Arbustini, E. 2010, "Mitochondrial DNA variant discovery and evaluation in human Cardiomyopathies through next-generation sequencing", *PloS one,* vol. 5, no. 8, pp. e12295.

Zaumsegel, D., Rothschild, M.A. & Schneider, P.M. 2013, "A 21 marker insertion deletion polymorphism panel to study biogeographic ancestry", *Forensic Science International-Genetics,* vol. 7, no. 2, pp. 305-312.

Zhang, S., Tian, H., Wang, Z., Zhao, S., Hu, Z., Li, C. & Ji, C. 2014, "Development of a new 26plex Y-STRs typing system for forensic application", *Forensic Science International: Genetics,* vol. 13, pp. 112-120.

Zhang, Y.D., Shen, C.M., Jin, R., Li, Y.N., Wang, B., Ma, L.X., Meng, H.T., Yan, J.W., Wang, D. & Yang, Z.L. 2015, "Forensic evaluation and population genetic study of

30 insertion/deletion polymorphisms in a Chinese Yi group", *Electrophoresis,* vol. 36, no. 9-10, pp. 1196-1201.

Zhang, Z., Wang, B.J., Guan, H.Y., Pang, H. & Xuan, J.F. 2009, "A LDR-PCR Approach for Multiplex Polymorphisms Genotyping of Severely Degraded DNA with Fragment Sizes < 100 bp", *Journal of forensic sciences,* vol. 54, no. 6, pp. 1304-1309.

Zhang, S., Sun, K., Bian, Y., Zhao, Q., Wang, Z., Ji, C. & Li, C. 2015, "Developmental validation of an X-Insertion/Deletion polymorphism panel and application in HAN population of China", *Scientific reports,* vol. 5, pp. 18336.

Zidkova, A., Horinek, A., Kebrdlova, V. & Korabecna, M. 2013, "Application of the new insertion-deletion polymorphism kit for forensic identification and parentage testing on the Czech population", *International journal of legal medicine,* vol. 127, no. 1, pp. 7-10.

Zimmermann, P., Vollack, K., Haak, B., Bretthauer, M., Jelinski, A., Kugler, M., Loidl, J. & Pflug, W. 2009, "Adaptation and evaluation of the PrepFiler™ DNA extraction technology in an automated forensic DNA analysis process with emphasis on DNA yield, inhibitor removal and contamination security", *Forensic Science International: Genetics Supplement Series,* vol. 2, no. 1, pp. 62-63.

Zipper, H., Brunner, H., Bernhagen, J. & Vitzthum, F. 2004, "Investigations on DNA intercalation and surface binding by SYBR Green I, its structure determination and methodological implications", *Nucleic acids research,* vol. 32, no. 12, pp. e103.

CHAPTER 9 APPENDIX

The following appendices are available in this section

Appendix I Table (A) – (E) on population differentiation

Appendix II Certificate of training of the researcher on COSHH Assessment

Appendix III: Ethical approval letter from University of Central Lancashire, Health, Safety, and Ethic Committee

Appendix IV: Approval letter of sample collection from Department of Chemistry Malaysia

Appendix V: Published journal article by the researcher

Appendix A: Table (A) – (E) on population differentiation

Table A: Population differentiation between Malays and Chinese, Indian, Iban, Bidayuh and thirteen regional populations.

Indel	M vs C	M vs Ind	M vs Ib	M vs By	M vs HC	M vs SK	M vs Ur	M vs CS	M vs BC	M vs P	M vs Fin	M vs S	M vs Z	M vs X	M vs Af	M vs InSA	M vs Mx
HLD77	0.0168	0.1211	0.1534	0.8262	***0.0000***	***0.0000***	0.0250	0.0234	0.2633	***0.0008***	***0.0000***	**1.0000**	0.3381	0.8229	***0.0001***	0.1761	0.0252
HLD45	***0.0000***	***0.0003***	***0.0020***	***0.0000***	***0.0000***	***0.0000***	0.5105	0.5748	**1.0000**	0.0965	0.0848	0.0833	***0.0000***	***0.0001***	0.4165	***0.0005***	0.0059
HLD131	0.6125	0.4101	0.0435	0.4642	0.2433	0.7445	***0.0001***	0.0033	0.5500	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	0.1164	***0.0000***
HLD70	0.2510	0.3703	0.1442	0.8299	0.0703	0.1489	0.5541	0.1818	0.4592	0.0041	***0.0000***	0.0068	***0.0000***	***0.0000***	***0.0008***	0.4275	0.2342
HLD6	0.2380	0.1766	**1.0000**	0.3130	0.8702	0.1523	0.0534	0.8303	0.4197	0.5470	0.0566	***0.0000***	***0.0000***	***0.0007***	0.5533	0.6847	0.0147
HLD111	***0.0001***	0.2100	0.0041	0.0154	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0014***	0.1154	***0.0000***	***0.0000***	0.0637	***0.0006***
HLD58	0.6213	0.1104	0.8416	***0.0001***	0.3887	0.0112	0.6322	0.0500	***0.0004***	0.3759	0.0526	0.0223	***0.0000***	***0.0000***	0.0583	0.0263	0.0828
HLD56	0.2332	***0.0007***	0.6750	**1.0000**	0.1389	0.1829	0.7123	***0.0013***	0.6336	***0.0022***	0.5060	***0.0003***	0.0745	**1.0000**	0.1765	***0.0000***	0.3640
HLD118	***0.0007***	***0.0000***	0.5232	0.0395	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***
HLD92	0.0730	0.1556	0.3185	0.8430	0.0090	0.0110	0.0152	0.5141	0.0640	0.1961	0.0197	***0.0027***	0.0181	0.0120	0.3886	0.4673	0.0297
HLD93	***0.0003***	0.4147	***0.0000***	***0.0000***	0.0029	***0.0000***	***0.0011***	0.0959	0.2537	0.2244	***0.0000***	0.1073	0.0132	0.0778	0.0486	0.3806	***0.0004***
HLD99	***0.0013***	0.1488	0.0102	0.4678	***0.0000***	***0.0000***	0.0028	0.0030	0.0295	***0.0002***	0.9150	0.8436	0.1831	0.4833	***0.0000***	0.0047	0.2207
HLD88	0.6866	0.0436	0.0413	***0.0012***	0.0525	0.5464	0.0089	0.3245	0.1961	0.6943	**1.0000**	***0.0000***	***0.0000***	***0.0000***	0.9209	***0.0024***	0.0053
HLD101	0.6802	0.2704	0.4200	0.9165	0.0620	0.5605	0.0324	0.0544	0.7240	0.3237	0.0080	***0.0000***	***0.0000***	***0.0000***	0.0131	0.1311	***0.0000***
HLD67	0.0045	0.1034	0.0377	0.2896	0.0667	0.3519	0.1151	0.7278	0.0031	0.7569	0.3539	***0.0000***	0.2829	**1.0000**	0.9138	***0.0018***	0.9201
HLD83	0.6937	***0.0015***	0.1468	**1.0000**	0.0438	0.2866	0.9254	0.1550	0.0497	0.1485	0.0820	0.0899	***0.0001***	0.0065	0.2019	0.0149	0.2140
HLD114	0.5341	0.0092	0.3644	0.4208	***0.0006***	0.0031	0.9253	0.1196	0.1889	0.6963	0.9239	***0.0000***	***0.0000***	***0.0000***	0.4750	0.8426	***0.0001***
HLD48	0.3683	0.1716	0.8271	0.4983	0.0127	0.0701	***0.0000***	***0.0000***	***0.0001***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0002***	***0.0000***
HLD124	0.3639	0.7562	0.2146	0.1618	0.5004	0.8671	0.1055	0.1406	0.0418	0.7701	0.8555	***0.0000***	***0.0000***	***0.0000***	0.3172	0.6208	0.0325
HLD122	0.2339	0.0053	***0.0011***	0.5941	0.0064	0.0010	***0.0000***	***0.0000***	***0.0004***	***0.0000***	***0.0000***	0.1022	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***
HLD125	0.2175	0.0722	0.6885	0.4389	***0.0014***	0.0075	0.9277	**1.0000**	0.9057	0.1305	0.3622	***0.0000***	***0.0000***	***0.0000***	0.7706	0.1097	0.5601
HLD64	0.6442	0.0446	**1.0000**	**1.0000**	0.1587	0.6411	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	0.2596	0.1447	***0.0000***	0.1646	0.0367
HLD81	0.0038	0.5092	0.9187	0.6469	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0003***	***0.0000***	0.0069	***0.0000***	***0.0023***	***0.0000***	***0.0000***	0.0406	0.0593
HLD136	0.7512	0.7654	***0.0001***	0.0065	0.6869	0.1833	0.4405	0.5912	***0.0022***	0.6991	0.3972	***0.0000***	***0.0012***	***0.0000***	0.3715	0.8518	0.8451
HLD133	0.3437	0.1144	0.5889	0.3532	0.2989	0.1110	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0001***	0.3482	0.1812	***0.0000***	0.2374	0.5963
HLD97	0.3455	0.5110	0.1402	0.4378	0.5424	0.8665	***0.0000***	***0.0005***	***0.0000***	***0.0000***	***0.0012***	0.5162	0.0154	0.3969	***0.0000***	0.7517	0.0522
HLD40	***0.0000***	***0.0002***	***0.0013***	***0.0000***	***0.0000***	***0.0000***	0.5072	0.5778	**1.0000**	0.0936	0.0766	0.0778	***0.0000***	***0.0000***	0.4282	***0.0002***	0.0129
HLD128	0.2147	***0.0001***	0.0397	0.0048	0.3907	0.2379	***0.0008***	0.0051	***0.0001***	0.0245	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0004***	***0.0000***
HLD39	0.0902	***0.0000***	0.3627	0.0034	0.4668	0.5598	***0.0000***	***0.0000***	0.0072	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***
HLD84	0.7227	0.0379	0.0397	***0.0016***	0.4609	0.8533	***0.0000***	***0.0018***	***0.0015***	0.0454	***0.0000***	0.0270	0.3193	0.4864	***0.0000***	0.1861	0.0430

Figures show p-value less than 0.05 (in bold and italic), figures show significant p-value after Bonferroni correction for 18 populations; 0.0028 (in bold, italic and highlighted in green). The exact test was carried out with 300,000 Markov steps. Figures in bold and highlighted in pink indicates no significant differentiation.

Table B: Population differentiation between Chinese and Indian, Iban, Bidayuh and thirteen regional populations.

Indel	C vs Ind	C vs Ib	C vs By	C vs HC	C vs SK	C vs Ur	C vs CS	C vs BC	C vs P	C vs Fin	C vs S	C vs Z	C vs X	C vs Af	C vs InSA	C vs Mx
HLD77	***0.0000***	0.3441	0.0356	0.0541	0.1273	0.7764	**1.0000**	0.3384	0.3802	0.0036	0.0030	0.1182	0.0055	0.0518	***0.0000***	0.7682
HLD45	***0.0000***	0.0786	0.8291	0.5486	0.0761	***0.0000***	***0.0000***	***0.0000***	***0.0011***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0001***	***0.0000***	***0.0000***
HLD131	0.8265	0.0109	0.1672	0.6530	0.7403	***0.0008***	0.0290	0.9051	***0.0002***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	0.3588	***0.0015***
HLD70	0.0331	0.8423	0.1251	0.7367	0.9295	0.5547	0.8215	0.8150	0.1177	0.0111	***0.0000***	***0.0000***	***0.0000***	0.0340	0.0399	0.0150
HLD6	0.0094	0.2554	0.9215	0.1655	0.8706	***0.0010***	0.1526	0.0495	0.5599	***0.0008***	0.0267	***0.0010***	0.0293	0.5562	0.4421	0.2263
HLD111	***0.0000***	0.3916	0.1797	0.4744	0.7625	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***
HLD58	0.0275	0.4284	***0.0000***	0.0973	***0.0010***	0.9300	0.1529	0.0038	0.7627	0.2086	0.0046	***0.0000***	***0.0000***	0.2038	0.0058	0.0201
HLD56	***0.0000***	0.4850	0.2390	0.9313	0.8693	0.3358	***0.0000***	0.1069	***0.0001***	0.0396	***0.0000***	0.6297	0.2152	0.0081	***0.0000***	0.0252
HLD118	***0.0000***	***0.0000***	0.1982	0.7438	0.5928	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***
HLD92	***0.0013***	0.4726	0.1343	0.7339	0.8730	0.7051	0.3248	0.9056	0.6201	0.7824	0.3198	0.6161	0.5581	0.3255	0.0090	0.7695
HLD93	0.0068	0.0285	0.4244	0.1280	0.8028	0.7763	0.1208	0.0613	0.0196	0.7823	0.0157	0.3241	0.0739	0.0943	0.0069	**1.0000**
HLD99	***0.0000***	0.5594	0.0165	**1.0000**	0.1701	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0001***	***0.0001***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***
HLD88	0.0112	0.1321	0.0039	0.0158	0.1885	0.0031	0.5829	0.3532	0.3962	0.5886	***0.0000***	***0.0000***	***0.0000***	0.7580	***0.0008***	***0.0013***
HLD101	0.5626	0.1810	0.8402	0.2456	**1.0000**	0.1193	0.1596	**1.0000**	0.6194	0.0302	***0.0000***	***0.0000***	***0.0000***	0.0635	0.3197	***0.0000***
HLD67	***0.0000***	0.4856	0.0874	0.0919	0.0075	***0.0000***	0.0030	***0.0000***	0.0008	***0.0000***	***0.0000***	0.0739	0.0050	***0.0014***	***0.0000***	0.0077
HLD83	0.0035	0.3606	0.7553	0.1649	0.6858	0.4947	0.0642	0.0115	0.0464	0.0207	0.0212	***0.0000***	***0.0016***	0.0691	0.0554	0.4840
HLD114	***0.0015***	0.8390	0.1273	0.0084	0.0300	0.3756	0.0231	0.5437	0.2680	0.5679	***0.0000***	***0.0000***	***0.0000***	**1.0000**	0.3496	***0.0000***
HLD48	0.0192	0.5739	0.9110	***0.0001***	0.0054	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***
HLD124	0.2104	0.8437	0.6881	0.0623	0.2956	0.0071	0.0166	***0.0025***	0.1777	0.4158	***0.0000***	***0.0000***	***0.0000***	0.0429	0.1288	0.2797
HLD122	0.1269	***0.0000***	0.6138	0.2687	0.0744	***0.0000***	***0.0000***	0.0128	***0.0004***	***0.0000***	0.8236	0.0120	***0.0003***	***0.0000***	***0.0027***	0.0031
HLD125	***0.0013***	0.1001	0.0367	0.0784	0.2606	0.2076	0.2638	0.2043	0.0050	0.0237	***0.0000***	***0.0000***	***0.0000***	0.1100	0.0026	0.5369
HLD64	0.0110	0.5374	0.5464	0.0319	0.2397	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	0.0989	0.0346	***0.0000***	0.0456	0.0060
HLD81	***0.0003***	***0.0016***	0.0155	0.6557	0.2402	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***
HLD136	0.4857	***0.0022***	0.0230	0.3524	0.0573	0.2496	0.3553	***0.0008***	0.4304	0.6607	***0.0000***	***0.0001***	***0.0000***	0.6991	0.9201	0.5449
HLD133	0.6045	0.7494	0.0510	0.8630	0.7340	***0.0001***	***0.0000***	***0.0000***	***0.0009***	***0.0000***	0.0266	0.9204	0.7619	***0.0000***	0.9168	0.6856
HLD97	0.0891	0.0101	0.0706	0.5443	0.3226	***0.0020***	0.0045	***0.0000***	0.0148	0.0197	0.7102	0.1787	0.9196	***0.0014***	0.1689	0.4130
HLD40	***0.0000***	0.0661	0.8215	0.5330	0.0723	***0.0000***	***0.0000***	***0.0000***	***0.0001***	***0.0000***	***0.0002***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***
HLD128	***0.0000***	0.4765	0.1262	0.0147	0.0031	***0.0000***	***0.0000***	***0.0000***	***0.0002***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***
HLD39	***0.0000***	0.4989	0.3279	0.1544	0.1028	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***
HLD84	0.0140	0.1179	0.0072	0.1893	0.4477	***0.0000***	***0.0002***	***0.0003***	0.0104	***0.0000***	0.0945	0.1307	0.2053	***0.0000***	0.0655	0.0129

Figures show p-value less than 0.05 (in bold and italic), figures show significant p-value after Bonferroni correction for 18 populations; 0.0028 (in bold, italic and highlighted in green). The exact test was carried out with 300,000 Markov steps. Figures in bold and highlighted in pink indicates no significant differentiation.

Table C: Population differentiation between Indian and Iban, Bidayuh and thirteen regional populations.

Indel	Ind vs Ib	Ind vs By	Ind vs HC	Ind vs SK	Ind vs Ur	Ind vs CS	Ind vs BC	Ind vs P	Ind vs Fin	Ind vs S	Ind vs Z	Ind vs X	Ind vs Af	Ind vs InSA	Ind vs Mx
HLD77	***0.0020***	0.0618	***0.0000***	***0.0000***	***0.0000***	***0.0000***	0.0128	***0.0000***	***0.0000***	0.0802	0.0101	0.1639	***0.0000***	0.8134	***0.0002***
HLD45	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	0.0090	***0.0012***	***0.0000***	0.0313	***0.0000***	0.1545	0.6404	***0.0000***	1.0000	0.2837
HLD131	0.0029	0.0891	1.0000	0.4585	0.0040	0.0497	1.0000	***0.0005***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	0.5425	***0.0021***
HLD70	0.0147	0.5811	***0.0008***	0.0095	0.1031	0.0167	0.1316	***0.0001***	***0.0000***	0.1050	***0.0000***	***0.0000***	***0.0000***	0.9077	0.8275
HLD6	0.1525	0.0196	0.0575	***0.0017***	0.6388	0.3816	0.8157	0.0356	0.7169	***0.0000***	***0.0000***	***0.0000***	0.0360	0.0613	***0.0000***
HLD111	***0.0000***	***0.0004***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	0.0640	0.8256	0.0137	***0.0000***	0.6717	0.0257
HLD58	0.1925	0.0427	0.2368	0.7373	0.0234	***0.0002***	***0.0000***	0.0059	***0.0003***	0.7024	***0.0000***	***0.0000***	***0.0006***	0.6057	1.0000
HLD56	***0.0005***	***0.0013***	***0.0000***	***0.0000***	***0.0000***	0.9009	0.0150	0.7353	***0.0020***	0.9152	***0.0000***	***0.0004***	0.0586	0.3044	0.0128
HLD118	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0023***	0.0245	0.0440	0.0046	***0.0000***	0.5977	***0.0015***	0.0046	0.0325	0.6114	0.0630
HLD92	0.0127	0.0931	***0.0000***	***0.0000***	***0.0003***	0.0467	0.0036	0.0032	***0.0002***	***0.0000***	***0.0000***	***0.0000***	0.0135	0.4801	***0.0000***
HLD93	***0.0000***	***0.0003***	0.0805	***0.0003***	0.0096	0.4499	0.7330	0.7644	***0.0010***	0.5197	0.1039	0.3733	0.2567	1.0000	0.0103
HLD99	***0.0001***	0.0189	***0.0000***	***0.0000***	0.2013	0.1355	0.4046	0.0214	0.1518	0.1399	0.9154	0.4614	***0.0001***	0.1823	0.8328
HLD88	***0.0000***	***0.0000***	0.6194	0.0510	0.8497	0.0038	***0.0025***	0.0915	0.0243	***0.0000***	***0.0004***	0.0386	0.0209	0.4180	0.4623
HLD101	0.0448	0.3705	0.7583	0.3524	0.4128	0.4446	0.5572	0.9257	0.1810	***0.0000***	***0.0000***	***0.0000***	0.2852	0.7684	***0.0001***
HLD67	***0.0001***	0.0067	***0.0002***	***0.0004***	0.8501	0.2324	0.1059	0.1562	0.4124	0.0036	0.0037	0.0785	0.1303	0.1058	0.0471
HLD83	0.0655	***0.0015***	0.0469	***0.0010***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	0.3866	0.0334
HLD114	***0.0001***	0.0920	***0.0000***	***0.0000***	0.0065	0.3621	***0.0006***	0.0218	0.0035	0.0209	***0.0000***	***0.0000***	***0.0006***	0.0141	0.2004
HLD48	0.1031	0.0349	0.5189	0.9341	***0.0015***	***0.0014***	0.0088	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	0.0203	***0.0001***
HLD124	0.1088	0.0685	0.8626	0.5285	0.2497	0.2757	0.0967	1.0000	0.5195	***0.0000***	***0.0000***	***0.0000***	0.5316	0.9189	0.0111
HLD122	***0.0000***	0.0321	0.4470	0.8515	***0.0003***	0.0167	0.2422	0.0221	***0.0000***	0.1153	0.3262	0.0427	***0.0003***	0.1591	0.1987
HLD125	0.1925	0.3560	***0.0000***	***0.0000***	0.0279	0.1019	0.1630	0.7682	0.2664	***0.0000***	***0.0000***	***0.0000***	0.1159	0.8464	0.0117
HLD64	0.0648	0.0595	0.2968	0.0312	0.0084	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0013***	0.3095	0.6116	***0.0000***	0.5376	1.0000
HLD81	0.6633	0.2237	***0.0000***	***0.0000***	***0.0000***	***0.0000***	0.0062	***0.0000***	0.0564	***0.0000***	0.0188	***0.0020***	***0.0000***	0.2209	0.2390
HLD136	***0.0000***	***0.0016***	1.0000	0.3694	0.7748	0.9110	0.0051	1.0000	0.2265	***0.0002***	***0.0011***	***0.0000***	0.1882	0.5408	0.9211
HLD133	0.3501	0.0085	0.3384	0.7505	***0.0007***	***0.0000***	***0.0002***	0.0058	***0.0002***	0.1263	0.4639	0.8388	***0.0004***	0.6822	0.3092
HLD97	0.5022	1.0000	0.1397	0.2332	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	0.1300	***0.0023***	0.1101	***0.0000***	0.7358	0.0067
HLD40	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	0.0055	***0.0027***	***0.0000***	0.0311	***0.0000***	0.1567	0.6505	***0.0000***	1.0000	0.2867
HLD128	***0.0000***	***0.0000***	***0.0000***	***0.0000***	0.0895	0.1524	0.7341	0.0124	0.1851	0.2457	***0.0022***	0.1516	0.8412	0.2789	1.0000
HLD39	***0.0000***	***0.0000***	***0.0000***	***0.0000***	0.6233	0.7336	0.0108	0.6909	0.0148	0.4758	***0.0000***	***0.0000***	0.9207	0.1650	0.7628
HLD84	***0.0000***	***0.0000***	0.0747	0.0195	0.0072	0.2641	0.1180	1.0000	***0.0013***	***0.0000***	0.3298	0.1963	0.0244	0.5160	1.0000

Figures show p-value less than 0.05 (in bold and italic), figures show significant p-value after Bonferroni correction for 18 populations; 0.0028 (in bold, italic and highlighted in green). The exact test was carried out with 300,000 Markov steps. Figures in bold and highlighted in pink indicates no significant differentiation.

Table D: Population differentiation between Iban and Bidayuh and thirteen regional populations.

Indel	Ib vs By	Ib vs HC	Ib vs SK	Ib vs Ur	Ib vs CS	Ib vs BC	Ib vs P	Ib vs Fin	Ib vs S	Ib vs Z	Ib vs X	Ib vs Af	Ib vs InSA	Ib vs Mx
HLD77	0.2920	***0.0013***	0.0064	0.4474	0.3699	0.8993	0.0599	***0.0002***	0.0759	0.6052	0.0786	***0.0025***	0.0051	0.4687
HLD45	0.1352	0.1064	0.6217	0.0075	***0.0001***	0.0028	0.1190	***0.0000***	0.0617	***0.0000***	***0.0000***	0.0113	***0.0000***	***0.0000***
HLD131	0.2590	***0.0001***	0.0040	***0.0000***	***0.0000***	0.0182	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0001***	***0.0000***
HLD70	0.0662	**1.0000**	0.6170	0.3363	**1.0000**	0.5581	0.2083	0.0284	***0.0000***	***0.0000***	***0.0000***	0.0676	0.0189	0.0061
HLD6	0.3569	**1.0000**	0.2022	0.0434	0.7451	0.3806	0.6303	0.0517	***0.0005***	***0.0000***	***0.0005***	0.6178	0.7682	0.0164
HLD111	0.7560	0.6857	0.1049	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***
HLD58	***0.0001***	0.6202	0.0394	0.4420	0.0177	***0.0002***	0.2348	0.0278	0.0428	***0.0000***	***0.0000***	0.0286	0.0520	0.1633
HLD56	0.6933	0.4170	0.4298	0.9253	***0.0002***	0.3556	***0.0003***	0.2249	***0.0000***	0.2076	0.6927	0.0743	***0.0000***	0.1493
HLD118	0.0041	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***
HLD92	0.4816	0.1835	0.2061	0.2234	0.8279	0.4289	0.8435	0.2358	0.0647	0.1867	0.1613	0.8362	0.0790	0.2754
HLD93	0.1824	***0.0000***	0.0104	0.0037	***0.0004***	***0.0000***	***0.0000***	0.0269	***0.0000***	***0.0014***	***0.0000***	***0.0001***	***0.0000***	0.0188
HLD99	0.0856	0.3967	0.0158	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0029***	***0.0013***	***0.0001***	***0.0003***	***0.0000***	***0.0000***	***0.0001***
HLD88	0.2090	***0.0000***	***0.0008***	***0.0000***	0.4453	0.7262	0.0137	0.0214	***0.0000***	***0.0000***	***0.0000***	0.0557	***0.0000***	***0.0000***
HLD101	0.2896	0.0039	0.0881	***0.0013***	0.0071	0.2500	0.0442	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0003***	0.0169	***0.0000***
HLD67	0.3640	0.4709	0.0945	***0.0001***	0.0260	***0.0000***	0.0212	***0.0001***	***0.0000***	0.3352	0.0451	0.0201	***0.0000***	0.0547
HLD83	0.1816	0.8594	0.4527	0.0913	0.0045	***0.0010***	0.0039	***0.0005***	***0.0013***	***0.0000***	***0.0000***	0.0057	0.3448	0.7668
HLD114	0.0669	0.0220	0.1096	0.2304	0.0142	0.7221	0.1556	0.3527	***0.0000***	***0.0000***	***0.0000***	0.8297	0.2154	***0.0000***
HLD48	0.7507	0.0070	0.0326	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***
HLD124	0.9194	0.0215	0.1809	0.0034	0.0063	***0.0013***	0.0953	0.2672	***0.0000***	***0.0000***	***0.0000***	0.0252	0.0687	0.4127
HLD122	***0.0002***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***
HLD125	0.7615	***0.0000***	***0.0009***	0.5127	0.6668	0.8168	0.3182	0.7809	***0.0000***	***0.0000***	***0.0000***	0.9227	0.2905	0.2973
HLD64	**1.0000**	0.1872	0.8226	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	0.3419	0.1882	***0.0000***	0.2306	0.0473
HLD81	0.4975	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0008***	***0.0000***	0.0132	***0.0000***	0.0039	***0.0005***	***0.0000***	0.0698	0.0863
HLD136	0.5331	***0.0000***	***0.0000***	***0.0000***	***0.0002***	***0.0000***	***0.0000***	0.0031	***0.0000***	***0.0000***	***0.0000***	0.0081	***0.0009***	***0.0000***
HLD133	0.1248	0.7950	0.3737	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	0.0076	0.8386	0.4723	***0.0000***	0.6085	**1.0000**
HLD97	0.5479	0.0156	0.0296	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	0.0161	***0.0004***	0.0190	***0.0000***	0.2801	***0.0005***
HLD40	0.1485	0.1025	0.6229	0.0077	***0.0002***	0.0030	0.1039	***0.0000***	0.0653	***0.0000***	***0.0000***	0.0187	***0.0000***	***0.0000***
HLD128	0.5242	***0.0012***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***
HLD39	0.0715	0.7078	0.5416	***0.0000***	***0.0000***	***0.0002***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***
HLD84	0.2837	***0.0008***	0.0028	***0.0000***	***0.0000***	***0.0000***	***0.0002***	***0.0000***	0.9014	***0.0024***	0.0034	***0.0000***	***0.0004***	***0.0001***

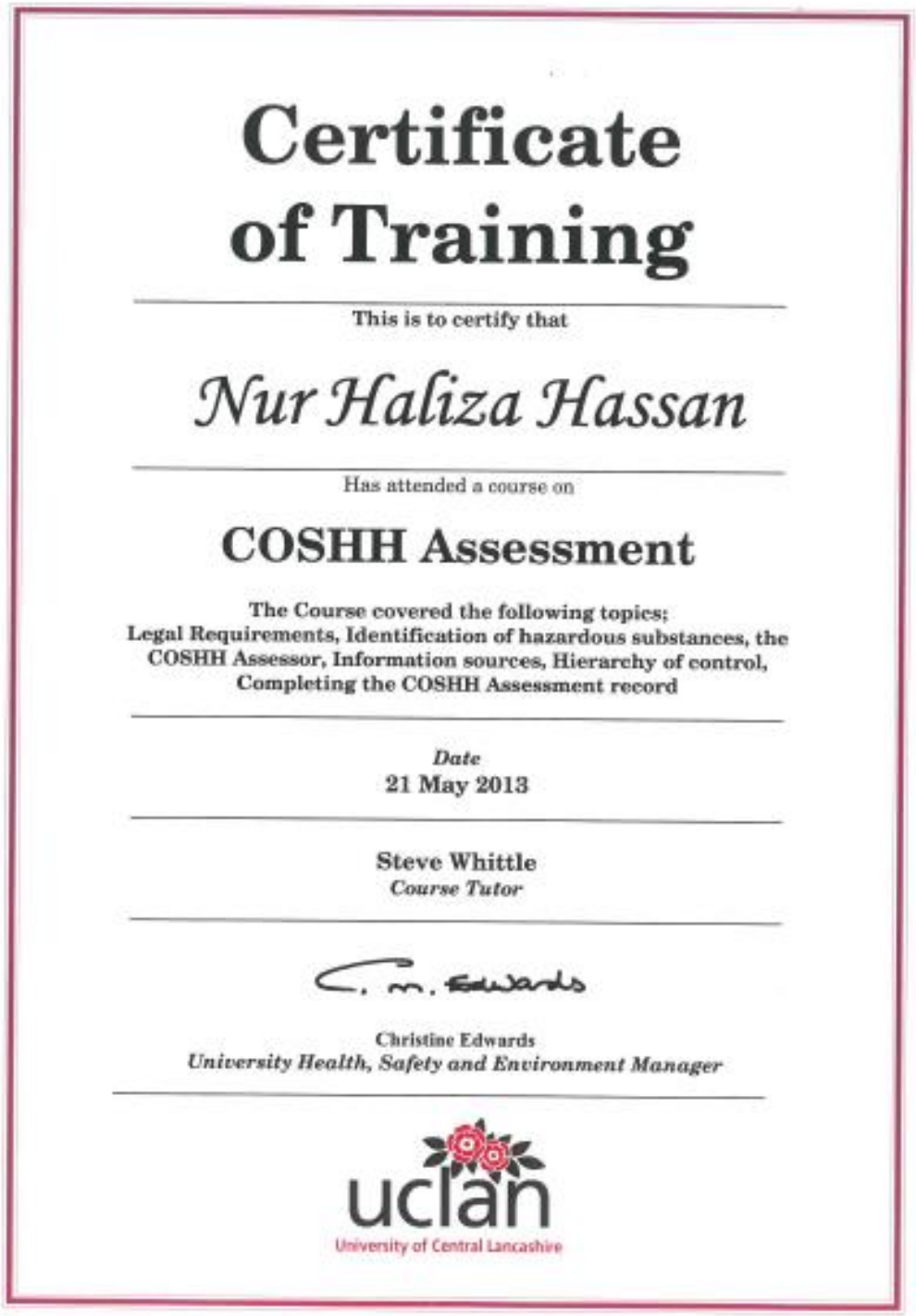
Figures show p-value less than 0.05 (in bold and italic), figures show significant p-value after Bonferroni correction for 18 populations; 0.0028 (in bold, italic and highlighted in green). The exact test was carried out with 300,000 Markov steps. Figures in bold and highlighted in pink indicates no significant differentiation

Table E: Population differentiation between Bidayuh and thirteen regional populations.

Indel	By vs HC	By vs SK	By vs Ur	By vs CS	By vs BC	By vs P	By vs Fin	By vs S	By vs Z	By vs X	By vs Af	By vs InSA	By vs Mx
HLD77	***0.0000***	***0.0000***	0.0494	0.0505	0.3822	***0.0009***	***0.0000***	0.6909	0.5168	0.5837	***0.0003***	0.0951	0.0555
HLD45	0.8556	0.1743	***0.0000***	***0.0000***	***0.0000***	***0.0014***	***0.0000***	***0.0004***	***0.0000***	***0.0000***	***0.0001***	***0.0000***	***0.0000***
HLD131	0.0262	0.1451	***0.0000***	***0.0002***	0.1792	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	0.0221	***0.0000***
HLD70	0.0231	0.0787	0.3151	0.0772	0.3330	***0.0004***	***0.0000***	0.0196	***0.0000***	***0.0000***	***0.0003***	0.6602	0.3796
HLD6	0.2250	0.9324	***0.0022***	0.2298	0.0813	0.6921	0.0035	0.0137	***0.0005***	0.0183	0.6932	0.5522	0.1605
HLD111	0.2858	0.0172	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0001***	***0.0000***	***0.0000***	***0.0000***	***0.0000***
HLD58	***0.0005***	0.0247	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	0.0527	***0.0014***	0.0039	***0.0000***	0.1532	0.0354
HLD56	0.1486	0.1799	0.7054	***0.0015***	0.6316	0.0037	0.5292	***0.0002***	0.0788	**1.0000**	0.1838	***0.0000***	0.3526
HLD118	0.1834	0.0130	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***
HLD92	0.0231	0.0252	0.0392	0.7331	0.1360	0.3166	0.0368	0.0078	0.0369	0.0305	0.6221	0.3152	0.0605
HLD93	0.0074	0.4111	0.2165	0.0116	0.0069	0.0014	0.5126	***0.0006***	0.0596	0.0080	0.0118	***0.0003***	0.3643
HLD99	***0.0016***	***0.0000***	***0.0000***	***0.0004***	0.0049	***0.0000***	0.2789	0.2376	0.0307	0.1424	***0.0000***	***0.0000***	0.0408
HLD88	***0.0000***	***0.0000***	***0.0000***	0.0407	0.1137	***0.0002***	***0.0003***	***0.0000***	***0.0000***	***0.0000***	***0.0008***	***0.0000***	***0.0000***
HLD101	0.1211	0.7481	0.0586	0.1087	0.8208	0.4359	0.0117	***0.0000***	***0.0000***	***0.0000***	0.0289	0.1929	***0.0000***
HLD67	0.6484	0.6818	0.0089	0.1696	***0.0000***	0.1507	0.0204	***0.0000***	**1.0000**	0.2947	0.1947	***0.0000***	0.3346
HLD83	0.0605	0.3379	0.8549	0.1221	0.0323	0.1237	0.0728	0.0786	***0.0002***	0.0053	0.1638	0.0187	0.2626
HLD114	***0.0000***	***0.0002***	0.4424	0.5063	0.0343	0.6872	0.2672	***0.0000***	***0.0000***	***0.0000***	0.0984	0.5353	0.0032
HLD48	***0.0007***	0.0078	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***
HLD124	0.0129	0.0982	***0.0005***	0.0046	***0.0014***	0.0555	0.1712	***0.0000***	***0.0000***	***0.0000***	0.0123	0.0441	0.5670
HLD122	0.0596	0.0098	***0.0000***	***0.0000***	***0.0020***	***0.0001***	***0.0000***	0.3723	***0.0011***	***0.0000***	***0.0000***	***0.0002***	***0.0002***
HLD125	***0.0000***	***0.0001***	0.3086	0.4337	0.5628	0.5447	**1.0000**	***0.0000***	***0.0000***	***0.0000***	0.6212	0.4756	0.1435
HLD64	0.1854	0.8146	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	0.3384	0.1808	***0.0000***	0.2137	0.0537
HLD81	***0.0004***	***0.0000***	***0.0000***	***0.0000***	***0.0001***	***0.0000***	***0.0015***	***0.0000***	***0.0003***	***0.0000***	***0.0000***	0.0118	0.0142
HLD136	***0.0000***	***0.0000***	***0.0004***	***0.0026***	***0.0000***	***0.0013***	0.0270	***0.0000***	***0.0000***	***0.0000***	0.0552	0.0072	***0.0012***
HLD133	0.0200	0.0037	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	0.0520	0.0146	***0.0000***	0.0314	0.1324
HLD97	0.1020	0.2002	***0.0000***	***0.0001***	***0.0000***	***0.0000***	***0.0000***	0.0845	***0.0010***	0.0820	***0.0000***	0.6606	0.0042
HLD40	0.8565	0.1869	***0.0000***	***0.0000***	***0.0000***	***0.0021***	***0.0000***	***0.0004***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***
HLD128	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***
HLD39	0.0061	0.0039	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***
HLD84	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	0.1519	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***

Figures above show p-value less than 0.05 (in bold and italic), figures show significant p-value after Bonferroni correction for 18 populations; 0.0028 (in bold, italic and highlighted in green). The exact test was carried out with 300,000 Markov steps. Figures in bold and highlighted in pink indicates no significant differentiation.

Appendix B: Certificate of training of the researcher on COSHH Assessment



Certificate of Training

This is to certify that

Nur Haliza Hassan

Has attended a course on

COSHH Assessment

The Course covered the following topics;
Legal Requirements, Identification of hazardous substances, the COSHH Assessor, Information sources, Hierarchy of control, Completing the COSHH Assessment record

Date
21 May 2013

Steve Whittle
Course Tutor

C. M. Edwards

Christine Edwards
University Health, Safety and Environment Manager

uclan
University of Central Lancashire

Appendix C: Ethical approval letter from University of Central Lancashire, Health, Safety, and Ethic Committee

30 September 2013

Will Goodwin / Nur Hassan
School of Forensic & Investigative Sciences
University of Central Lancashire

Dear Will / Nur

Re: STEM Ethics Committee Application
Unique Reference Number: STEM 148

The STEM ethics committee has granted approval of your proposal application '**Evaluation of Insertion/Deletion Polymorphisms (INDELS) in the Malaysian Populations and Forensic Casework Studies**'.

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 quoting your unique reference number.

Yours sincerely

Kevin Butt
Vice Chair
STEM Ethics Committee

Appendix D: Approval letter of sample collection from Department of Chemistry Malaysia

JABATAN KIMIA MALAYSIA
KEMENTERIAN SAINS, TEKNOLOGI DAN INOVASI
DEPARTMENT OF CHEMISTRY MALAYSIA
MINISTRY OF SCIENCE, TECHNOLOGY AND INNOVATION

Our Ref. : JK/10/FOR/500-7/1/2 Jld.2 (18)
Date : 29th May 2013

Mrs. Nur Haliza Hassan
PhD in Forensic DNA
School of Forensic and Investigative Sciences
University of Central Lancashire
PR1 2HE Preston
United Kingdom

Dear Mrs. Nur Haliza,

REQUEST FOR SPECIMENS FROM THE DEPARTMENT OF CHEMISTRY MALAYSIA

With regards to your letter dated 28th May 2013, Requesting For Specimens From The Department of Chemistry Malaysia, for your PhD research project.

The Department support your application based on the following conditions :

(i) Availability of the samples;
(ii) Compliance with all the necessary ethical requirements;
(iii) Only for research purposes;
(iv) No information regarding the identity of the samples and genetic deficiency of a particular population / ethnic group shall not be revealed/reported; and
(v) The samples shall be collected by Nur Haliza from the Department of Chemistry Malaysia.

Thank You.

Yours sincerely,

(LIM KONG BOON)
Director of Forensik Division
Department of Chemistry Malaysia
☎ : 603-7985 3870
✉ : kblim@kimia.gov.my

c.c. **Dr. William Goodwin**
School of Forensic and Investigative Sciences
University of Central Lancashire
PR1 2HE Preston
United Kingdom

Jalan Sultan, 46661 Petaling Jaya, Selangor Darul Ehsan
Tel.: 603 - 7985 3000 Fax: 603 - 7985 3092
www.kimia.gov.my

Appendix E: Published journal article by the researcher

Forensic Science International: Genetics Supplement Series 5 (2015) e207–e209

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

journal homepage: www.elsevier.com/locate/FSIGSS

Collection protocols for the recovery of biological samples

D. Aloraer*, N.H. Hassan, B. Albarzinji, W. Goodwin

School of Forensic and Applied Sciences, University of Central Lancashire, Preston, PR1 2HE, United Kingdom

ARTICLE INFO

Article history:
Received 3 September 2015
Accepted 15 September 2015
Available online 24 September 2015

Keywords:
Forensic biology
DNA collection
Double swabbing

ABSTRACT

The main focus in forensic genetics for two decades has been to improve the extraction of DNA from a wide variety of evidence and to make the profiling technology more sensitive and robust. In contrast, the recovery methods for biological material have seen little development. This study aims to improve the efficacy of the collection and storage processes, from crime scene to receipt at the laboratory. This study compared the use of ultrapure water as a wetting agent when collecting biological evidence using swabs with a detergent-based buffer. The results show that the stability post-collection greatly improved by using a newly developed buffer. When ultrapure water is used, DNA degradation was seen after 6 h at room temperature. However, the detergent-based buffer stabilized DNA for up to 48 h, even when the temperature was increased to 50 °C. The impact of these findings may be limited where crime scene evidence can be refrigerated until it reaches the laboratory. However, there are many situations/contexts where sample refrigeration is not possible and there is scope to improve the preservation of the generic forensic evidence before it reaches the laboratory.

1. Introduction

When analyzing biological evidence the precursors to extraction and analysis are the sample collection, handling and storage. Collection, preservation and storage of biological evidence have a fundamental impact on the quality of the sample and the resultant DNA profile. If the sample quality is damaged by poor practices it can undermine the potential for it to be used as evidence [1]. These initial steps must be undertaken carefully, and the most reliable and reproducible protocols should be used for the crime scene collection and preservation of the biological materials before they reach the laboratory. DNA collection procedures can vary, but often start at the crime scene where biological samples such as blood, semen and saliva are identified, collected, and then transferred to the forensic laboratory. Many samples are collected at crime scenes using DNA-free water as a wetting agent with cotton swabs—this basic system does not have any preservative properties.

2. Material and method

Saliva samples were deposited on a plastic, metal and glass substrates and allowed to dry overnight and then recovered using the double swab technique [2,3] using Ultrapure water and a detergent-based wetting agent (PrestGene, UK). PureGene extraction (Qiagen) was used after the samples had been exposed (post-collection) to various environmental conditions (−20 °C, room temperature, 37 °C and 50 °C). The quantity of extracted DNA present in samples was measured using real-time PCR, whereas the quality of the extracted DNA present was determined using in-house multiplex PCR.

2.1. DNA quantification

All extracted DNA samples from Saliva were quantified using the Quantifiler™ Human DNA Quantification kit on the ABI7500 real-time PCR machine (Applied Biosystems, UK). Amplification reactions and amounts were carried as recommended by the manufacturer.

2.2. DNA Amplification

DNA (1 μL) was amplified using the polymerase chain reaction (PCR) using an in-house assay that amplifies four amplicons ranging between 50 bp and 154 bp.

2.3. DNA analysis

DNA fragment analysis was carried out on ABI 3500 Prism® Genetic Analyzer.

* Corresponding author.
E-mail address: Dlnaloaer@uclan.ac.uk (D. Aloraer).

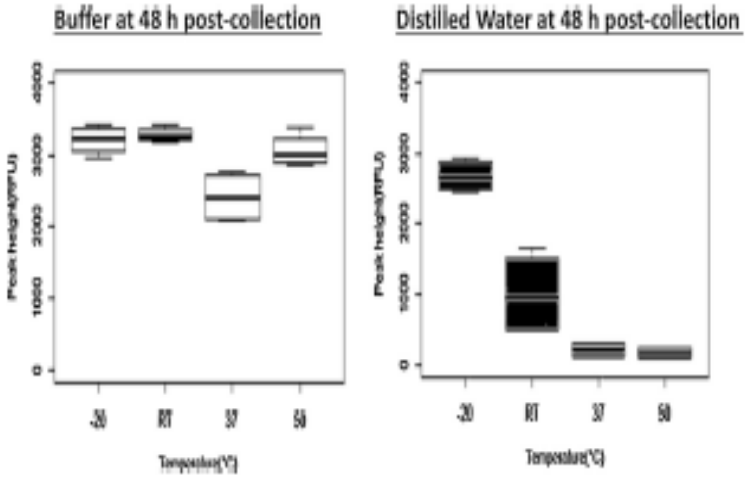


Fig. 1. Boxplots demonstrate the peak heights of Saliva samples collected by distilled water and buffer separately, at four different temperatures (−20 °C, room temperature (RT), 37 °C and 50 °C) for 48 h post-collection, and amplified by mini 4-plex multiplex PCR.

2.4. Data analysis

The data obtained from the capillary electrophoresis (CE) were analysed using an ABI3500 with GeneMapper® ID-X Software Version 1.2 (Life Technologies™, UK). The peak heights (RFU) of the samples were used to perform the statistical analysis. R-Studio software was used for statistical analyses such as the analysis of variances (ANOVA).

3. Results and discussion

Samples were recovered in triplicates from glass, plastic and metal, results below are of plastic only using the Double Swab technique [2,3] using ultrapure water and a detergent-based buffer and stored post-collection for up to 48 h at different temperatures (−20 °C, room temperature, 37 °C and 50 °C). The quantitation data is shown in Fig. 1 and the profiles generated using an in-house

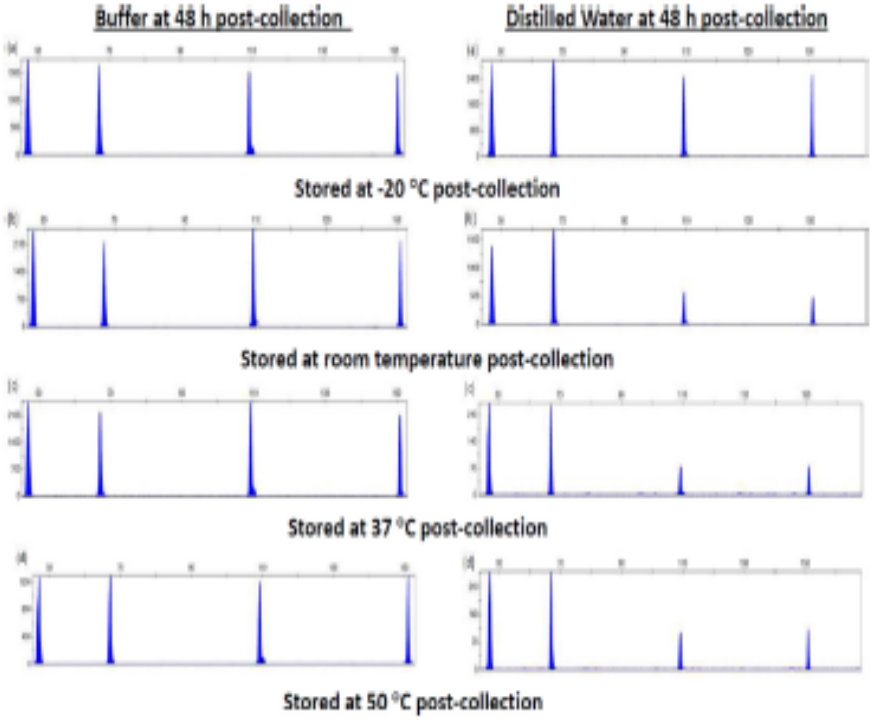


Fig. 2. Electropherograms show the mini 4-plex multiplex PCR amplification of extracted DNA saliva samples after 48 h post-collection by buffer against distilled water in different temperatures (−20 °C, room temperature, 37 °C and 50 °C).

multiplex are shown in Fig. 2. At 48 h there is a statistically significant difference between the amount of DNA using the two collection methods ($F(3, 12) = 50.43$, $P = 4.47e\text{-}07$)).

The data presented demonstrates that post-collection degradation of biological material could be a significant problem when working in environments were optimum storage of the swabbed material is not possible. Using a detergent-based buffer stabilized the DNA on the swabs, as illustrated both by the quantitation and the quality of the DNA profiles, both in terms of peak height and balance.

4. Conclusion

Further work need to be done to examine the efficacy of the detergent-based buffer when collecting biological evidence such as touch DNA. Gaining a better understanding of the key factors in crime scene sample collection and post-collection handling that impact on DNA recovery should give a clear indication of best practices in post-collection sample handling whilst in transit to the laboratory.

Conflict of interest

None.

References

[1] J. Bonnet, M. Colotte, D. Coudy, P. CouallierV, J. ortier, B. Morin, S. Tuffet, Chain and conformation stability of solid-state DNA: implications for room temperature storage, Nucleic Acids Res. 38 (5) (2010) 1531–1546.
[2] D. Sweet, M. Lorente, A. Valenzuela, J.A. Lorente, E. Villanueva, An improved method to recover saliva from human skin: the double swab technique, J. Forensic Sci. 2 (1997) 320–322.
[3] D. Sweet, M. Lorente, A. Valenzuela, J.A. Lorente, J.C. Alvarez, D.N.A. Increasing, Extraction yield from saliva stains with a modified Chelex method, Forensic Sci. Int. 83 (3) (1996) 167–177.1996.