Application of single nucleotide polymorphisms (SNPs) to forensic casework in Malaysia

by

Sharizah Alimat

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

March 2014

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:

___

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

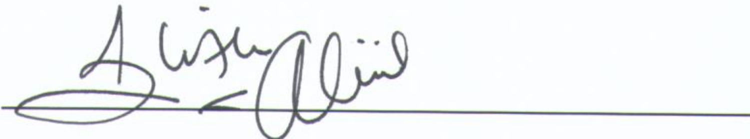
___

(state award and awarding body and list he 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 Doctor of Philosophy (PhD)

School Forensic and Investigative Sciences

ABSTRACT

The analysis of degraded DNA can be problematic. Recent advances in the identification and analysis of single nucleotide polymorphisms (SNPs) have demonstrated the advantage of these markers over short tandem repeats (STRs) in that they only require small amplicons. However, before applying to casework, it is important to develop allele frequency databases from relevant populations. The main aim of the study is to characterize three Malaysian major ethnic groups; Malay, Chinese and Indian, using 52 autosomal SNP markers that have been identified in the SNP*for*ID project.

Sanchez et. al., 2006 reported a multiplex of 52 SNP markers in one PCR reaction with two single base reaction (SBE) in the detection of SNPs using capillary electrophoresis (CE). The amplicons for PCR ranged from 59 bp to 115 bp. Whilst for SBE reactions ranged from 16 nt to 92 nt. In their study, full complete profile was obtained from 500 pg DNA input. The study was carried out on three major populations: African, Asian and European.

As in this study, a total of 325 Malaysian samples (109 from Malays, 107 from Chinese and 109 from Indians) were genotyped. In order to genotype the population samples reliably and robustly, four sets of 13-plex SNPs were developed. Internal validation was carried out using both genetic analyzers, ABI PRISM® 310 and 3500 Genetic Analyzer. Sensitivity and reproducibility studies demonstrated that the assays were highly sensitive, requiring as little as 30 pg of DNA. Full, complete and clear profiles were generated. Data were collected and evaluated statistically for forensic usefulness.

Across the three ethnic groups, few significant departures from HWE, at p values <0.05 were observed at 3 SNP markers in Malays, 4 SNP markers in Chinese and 9 SNP markers in Indian samples. Five markers (rs2107612, rs722098, rs2076848, rs907100 and rs1528460) in the Indians and one marker (rs1528460) in the Chinese, showed the lowest p value, that is $p=0.0$. However, after Bonferroni correction at $p <0.00096$ significant deviation(s) from HWE was observed at 1 SNP marker (code marker 26) in the Malays, 2

SNP markers (code marker 46 and 54) in the Chinese and 5 SNP markers (code marker 12, 21, 36, 38 and 54) in the Indians. In addition, a pair of loci (at code markers 3 and 53) was found to be associated in the Malays after the Bonferroni correction (at $p < 0.0000377$).

As for forensic parameters, the combined mean match probabilities for the 52 SNPs of Malay, Chinese, and Indian were 1 in $3.9654e^{-19}$, $5.3964e^{-18}$, and $1.7459e^{-19}$, corresponding to a combined power of discrimination of >99.99999999%, respectively. Paired F_{st} values obtained in the study showed, as expected, that Malay group is closely related to the Chinese population, with the Indian population being more distant.

CONTENTS

DECLARATION .... **i**

ABSTRACT .... **ii**

CONTENTS .... **iv**

LIST OF FIGURES .... **ix**

LIST OF TABLES .... **xvi**

ACKNOWLEDGEMENTS .... **xxi**

CHAPTER 1 .... **1**

INTRODUCTION **1**

1.1. GENETIC MARKERS IN THE FORENSIC DNA ANALYSIS 2

1.2. SINGLE NUCLEOTIDE POLYMORPHISMS (SNP) 5

1.2.1. SNP Characteristics 7

1.2.2. SNP Consortium/Databases 10

1.2.3. Categories of SNP Markers 16

1.2.4. Widely-used SNP Typing Methodologies 18

1.3. SNPS AS FORENSIC MARKERS 20

1.3.1. Degraded DNA 20

1.3.2. Human Identification 21

1.3.3. Estimating ethnic origin of a sample 23

1.4. APPLICATION OF SNPs MARKERS ON THE MALAYSIAN POPULATIONS AND FORENSIC CASEWORK 24

1.4.1. Geography and demographic of Malaysia 24

1.4.2. Multi-racial residents in Malaysia 26

1.4.3. Forensic Casework in Malaysia 26

1.5. STUDIES RELATED TO MALAYSIA 28

1.5.1. Autosomal SNP studies on Southeast Asian and East Asian populations 28

1.5.2. Theories on migration routes within Asia 32

1.6. OVERVIEW AND AIMS OF THE STUDY 35

CHAPTER 2 .... **38**

MATERIALS AND METHODS 38
2.1 Sampling...39
2.1.1. Malaysian population samples 39
2.1.2. Environmental Insult samples...39
2.1.3. Forensic casework samples..40
2.1.4. Kinship samples..40
2.2. Selection of Single Nucleotide Polymorphisms (SNP) markers 40
2.3. Quality Control..40
2.4. DNA extraction and purification 43
2.4.1. Organic Solvent (Phenol/Chloroform/Isoamyl alcohol)..43
2.4.2. QIAamp DNA Micro kit...43
2.4.3. QIAamp DNA Investigator Kit...44
2.5. Quantification of DNA samples...45
2.6. Reproducibility and Sensitivity study...46
2.6.1. Serial dilutions...46
2.6.2. Testing population samples 47
2.7. SNP Genotyping 47
2.7.1. TaqMan® SNP Genotyping Assays 47
2.7.2. SNaPshot® Multiplex System 51
2.8. Data collection and evaluation. 62
2.9. Statistical analysis 62
CHAPTER 3 64
DEVELOPMENT, OPTIMIZATION AND VALIDATION OF THE SNP DETECTION ASSAYS 64
PART ONE:... 65
***TAQMAN® SNP GENOTYPING ASSAYS* 65**
3.1. INTRODUCTION...65
3.2. RATIONALE OF THE STUDY...70
3.3. RESULTS 71
3.3.1. Sensitivity and reproducibility of the TaqMan® SNP Genotyping assays..................71
3.3.2 Preparation of Allelic Discrimination (AD) study..71
3.3.3. TaqMan® SNP Genotyping assays optimization 75
3.3.4. TaqMan® SNP Genotyping analysis on testing population samples..........................79

3.4. DISCUSSION....................80

3.4.1. Allelic discrimination using TaqMan® SNP Genotyping assays and AB® 7500 Real-Time PCR System....................80

3.4.2. Detection level of the TaqMan® SNP Genotyping assays on few parameters....................81

3.4.3. Limitation of TaqMan® SNP assays....................82

PART TWO:.................... 84

SNAPSHOT ASSAYS 84

3.1. INTRODUCTION....................84

3.2. RATIONALE OF THE STUDY....................86

3.2.1. Sensitivity and reproducibility evaluation of the SNaPshot assay(s)....................86

3.2.2. Developing a series of multiplex assays to be used in the study....................88

3.2.3. Optimization and validation of the selected SNaPshot multiplex assays....................88

3.2.4. SNP genotyping and characterization of 30 unrelated testing individuals....................88

3.3. RESULTS....................89

3.3.1. Single SNaPshot reaction....................89

3.3.2. Series of multiplexes developed throughout the study....................90

....................98

3.3.3. Development, optimization and validation of the 13-plex PCR/SNaPshot reactions; 13_{st}, 13_{nd}, 13_{rd} and $13_{th.}$....................98

3.3.4. Evaluation of 13-plex SNaPshot assays with testing population samples....................103

....................105

3.4. DISCUSSION....................107

3.4.1. Selection of SNP markers and platform....................107

3.4.2. Development of the multiplex assays....................108

3.4.3. The stability of the multiplex assays....................110

CHAPTER 4 112

VERIFICATION STUDIES: ANALYSIS OF FOUR 13-PLEX SNAPSHOT ASSAYS USING ABI PRISM® 3500 GENETIC ANALYZER 112

4.1. INTRODUCTION....................113

4.2. RATIONALE OF THE STUDY....................115

4.2.1. Concordance analysis using the 3500 Genetic Analyzer....................115

4.2.2. Assessment of the 13-plex assays using different polymers and parameters....................115

4.2.3. Analysis of tested samples with optimal parameters and conditions......................118
4.3. RESULTS 118
4.3.1. Separation and detection of 13-plex assays on the 3500 118
4.3.2. Effect of different polymers and parameters on the 13-plex assays analyzed using the 3500 Genetic Analyzer 120
4.3.3. Sensitivity and reproducibility assessment using polymer POP-6™ and optimal parameters on the 3500 Genetic Analyzer 125
4.4. DISCUSSION......134
4.4.1. Corroboration analysis with the 3500......134
4.4.2. DNA separation and resolution analysis with different polymer types 134
4.4.3. Overview of polymer solutions, includes POP-4™, POP-6™ and POP-7™......135
4.4.4. Verification studies under optimal conditions and parameters on the 3500 Genetic Analyzer 137
CHAPTER 5 140
SNP GENOTYPING OF MALAYSIAN POPULATION SAMPLES WITH 13-PLEX PCR/SBE ASSAYS...... 140
5.1. INTRODUCTION......141
5.1.1. 52-SNPforID plex 142
5.1.2. SNPforID Browser......143
......143
5.2. RATIONALE OF THE STUDY......144
5.2.1. Malaysian population samples 144
5.2.2. Data Collection and statistical evaluation 144
5.2.3. The Snipper app suite online software......144
5.2.4. Data comparison between Malaysia and neighbouring countries......146
5.3. RESULTS 148
5.3.1. Allele Frequencies 148
5.3.2. Tests for Hardy-Weinberg Equilibrium (HWE) 151
5.3.3. Exact test for Linkage Equilibrium (LE)......155
5.3.4. Population Genetic Structure 160
5.3.5. Population assignment of individual genotypes 160
5.3.6. Forensic statistical analysis......164

5.3.7. Comparison of forensic parameters between Malaysian SNP data with Malaysian STR data 168

5.5. DISCUSSION....169

5.5.1. Allele distribution among the populations....169

5.5.2. Hardy-Weinberg Equilibrium (HWE) and LE departures....169

5.5.3. Pairwise FST estimates between the populations....170

5.5.4. The relationship between the Malays, the Chinese and the Indians....171

CHAPTER 6 172

ANALYSIS OF 13-PLEX SNAPSHOT ASSAYS ON THE FORENSIC BIOLOGICAL SAMPLES... 172

6.1. INTRODUCTION....173

6.2. RATIONALE OF THE STUDY....174

6.2.1. Section One: Kinship Analysis....174

6.2.2. Section Two: Crime Samples....174

6.3. RESULTS....175

6.3.1. Section One: Kinship Analysis....175

....178

6.3.2. Section Two: Analysis of the crime samples....180

....188

6.4. DISCUSSION....190

6.4.1. Section One: Kinship studies....190

6.4.2. Section Two: The analysis of forensic biological samples....191

CHAPTER 7 193

GENERAL DISCUSSION AND FUTURE WORK.... 193

7.1. GENERAL DISCUSSION....194

Limitations/Challenges of SNP Analysis....197

7.2. FUTURE WORK....197

APPENDICES 198

REFERENCES 246

LIST OF FIGURES

Figure 1-1: Examples of mutations affecting a few base pairs, including examples of transition and transversion substitutions [Taken from Jobling et.al. 2013]……………………..…6

Figure 1-2: Transition and transversion mutations [Taken from Jobling et.al. 2013]…………7

Figure 1-3: Overview of mutation rates (μ) of different classes of polymorphisms [Taken from Jobling et.al. 2013]…………………………………………………………………………………………...8

Figure 1-4: Population density by state Malaysia, 2010 [Taken from Department of Statistics 2010]……………………………………………………………………………………………………...25

Figure 1-5: Numbers on the tree branches are bootstrap values. (A) Phylogeny of group private haplotypes. EA and SEA private haplotypes demonstrates sharing a common origin. Shared haplotypes are found in all EA, SEA, and CSA samples; African haplotypes were only used as out-group. (B) Maximum likelihood tree of 29 populations. The tree is based on data from 19,934 SNPs. [Taken from (Abdulla et al. 2009)]………………………………………29

Figure 1-6: Neighbour-joining tree on 15 populations based on the F_{ST} measurement. [Taken from (Hatin et al. 2011)]……………………………………………………………………………31

Figure 1-7: Colored arrows show the genetic diversification of humans after they migrated eastward along what is now India's coast and split into numerous genetically distinct groups that moved across Southeast Asia and migrated north into East Asia [Taken from (Normile 2009)]…………………………………………………………………………………………………….32

Figure 2-1: Overview of the procedure of PCR and SBE reactions for SNP genotyping whole process [Taken from Applied Biosystems 2009a]………………………………………………….51

Figure 3-1: Allelic discrimination is achieved by the selective annealing of TaqMan® MGB probes. [Taken from Applied Biosystems, 2010a]…………………………………………………….66

Figure 3-2: The circled sample in the scatter plot (A) is assigned an undetermined call when the default threshold of 0.95 is applied for analysis, but it is assigned a heterozygous call (B) when the Quality value threshold is lowered to 0.80. [Taken from Life Technologies (2013)]……68

Figure 3-3: Illustration of angle (A) and amplitude (B) calculated from the No Template control (NTC). Good cluster separation means good separation of data points from other clusters (angle) and leads to good angle scores. Good signal intensity means good separation of data points from the NTC and generates good amplitude scores. [Taken from Life Technologies (2013)]……………………………………………………………………………………………69

Figure 3-4: Distribution of angles (A) and amplitudes (B). The distribution of angles (A) shows three well separated clusters. The distribution of amplitudes (B) shows three clusters with good signal intensity that is well separated from the NTC. [Taken from Life Technologies (2013)]……………………………………………………………………………………………………69

Figure 3-5: (a) An amplification plot of the AQ analysis, and (b) Report of the AQ analysis; that showed the targeted SNP marker and cycle threshold (C_T) of the samples………………73

Figure 3-6: Detected alleles (homozygote A, homozygote G, heterozygote alleles AG and undetermined alleles) were plotted in Allelic Discrimination tab……………………………………74

Figure 3-7: Report generated from the Post-Read analysis………………………………………………75

Figure 3-8: Effects of DNA concentrations on Rn versus cycles………………………………………..75

Figure 3-9: Genotype clusters and allele calls generated from the optimization of various PCR reaction volumes and series of control DNA concentrations, (i) total PCR reaction of 5 µl, (ii) 10 µl, (iii) 12.5 µl and (iv) 25 µl…………………………………………………………………………77

Figure 3-10: Clusters of data generated from testing 19 population samples. All PCR reactions were in 5 µl, but with two different DNA concentrations; (a) 0.125 ng and (b) 0.0078 ng……78

Figure 3-11: Clustered plots of 90 Malaysian DNA samples and 3 NTCs using the optimal conditions for 2 SNP markers: (a) rs1493232 (A/C), and (b) rs1335873 (A/T)…………………..79

Figure 3-12: A 52 SNP markers assay with two sets of SBE reactions; A) 23 markers, and B) 29 markers. [Taken from the original study (Sanchez et al. 2006a)]………………………………..85

Figure 3-13: Single nucleotide primer extension [Taken from Applied Biosystems, 2010b]……….………………………………………………………………………………………………………87

Figure 3-14: Amplicons generated from the single reaction assay with the effects of primer concentrations…………………………………………………………………………………………………89

Figure 3-15: Three multiplex assays; 3-plex, 5-plex and 10-plex of TaqMan control DNA (0.5 ng). Arrows show the drop-out allele C (for code marker 10)…………………………………..92

Figure 3-16: Three different multiplex assays for the second trial in multiplexing the markers; 3-plex2, 5-plex2 and 10-plex2 of TaqMan control DNA (0.5 ng)………………………..94

Figure 3-17: An example of the first 13-plex assays (1st-plex, 2nd-plex, 3rd-plex and 4th-plex) of M7-7 DNA sample (0.5 ng) developed in this study………………………………………….….96

Figure 3-18: An example of the second 13-plex assays (1st2-plex, 2nd2-plex, 3rd2-plex and 4th2-plex) of M7-7 DNA sample (0.5 ng) developed in this study……………………………………98

Figure 3-19: Human control DNA with 1 µl of serial dilution (0.03 ng). The arrow shows an allele C that dropped-out…………………………………………………………………………………………99

Figure 3-20: Full SNP profile of human control DNA with 1 µl of serial dilution (0.125 ng)……..101

Figure 3-21: Full SNP profile of human control DNA with 1 µl of serial dilution (0.5 ng)…101

Figure 3-22: Full SNP profile of human control DNA with 1 µl of serial dilution (1.0 ng)…102

Figure 3-23: Three population samples (1 ng each) analyzed using the *13*$_{nd}$ bin created using the GeneMapper v 3.2……………………………………………………………………………………102

Figure 3-24: Full SNP profile of an individual with 0.04 ng DNA input……………………….……104

Figure 3-25: Partial SNP profile generated from undetermined DNA template……………..104

Figure 3-26: SNP profile from an individual with 1 µl (70 ng) of DNA template……….……105

Figure 3-27: SNP profile of an individual with 1 µl (569 ng) of DNA template………………..106

Figure 4-1: Illustration of internal capillary environment and forces with moving DNA molecules through a sieving polymer buffer. [Taken from (Butler 2011)]……………………..116

Figure 4-2: Electropherograms displaying the separation of: (A) SNaPshot profiles of *13*$_{th}$ multiplex assay of Control DNA, and (B) SNaPshot profiles of *13*$_{rd}$ multiplex assay of 3 different individual samples using POP-7™ on 3500 Genetic Analyzer…………………………119

Figure 4-3: Comparison of SNaPshot profiles of *13*$_{th}$ multiplex assay using polymers; A) POP-4™ on ABI 310 PRISM® Genetic Analyzer, B) POP-4™, C) POP-6™ and D) POP-7™ on the 3500 Genetic Analyzer. Boxes show the most affected SNPs alleles…………….………121

Figure 4-4: Alleles migration of Control DNA in different polymer types; A) POP-4™, B) POP-6™ and C) POP-7™ on 3500 Genetic Analyzer. Code markers' positions were highlighted in red. The arrows show the first allele's movement in each dye………………..122

Figure 4-5: (A) SNP profiles of Positive Control of ABI PRISM® SNaPshot Multiplex kit and (B) *13*$_{rd}$-plex profiles of an individual sample electrophoresed in different polymer types; POP-4™, POP-6™ and POP-7™ on 3500 Genetic Analyzer. Arrows show poor separation of the allele's occurrence……………………………………………………………………………………………..124

Figure 4-6: A) and B) are the profiles of *13*$_{nd}$–plex assay, but with 0.4 µl and 1 µl of samples loaded, respectively. C) and D) are the profiles of *13*$_{th}$-plex assay, with 0.4 µl and 1 µl of samples loaded, respectively using polymer POP-6™ on the 3500 Genetic Analyzer……..125

Figure 4-7: Peak height (RFU) comparison of three SNP electropherograms of the same Control DNA sample (0.5 ng/µl) using; (i) an optimal conditions and parameters with ABI 310 PRISM® Genetic Analyzer, (ii) Same product as (A) but electrophoresed and analyzed with 3500 Genetic Analyzer, and iii) new optimal conditions with 3500 Genetic Analyze……126

Figure 4-8: Electropherograms of; A) triplicates of *13$_{th}$*–plex assay with 2.5 ng of Control DNA and B) triplicates of *13$_{nd}$*-plex assay with 0.1 ng of Control DNA, using polymer POP-6™ and 3500 Genetic Analyzer, each at constant 9000 RFU and 6000 RFU, respectively……….128

Figure 4-9: *13$_{nd}$* -plex profiles of a series of DNA template input: all are shown with a scale of 4000 RFU. ………129

Figure 4-10: SNP electropherograms of *13$_{st}$*–plex assay generated from DNA serial dilutions: A) 0.01 ng, B) 0.02 ng, C) 0.03 ng and D) 0.06 ng. The arrows show the targeted allele(s) was dropped-out……………………………………………………………………………………………130

Figure 4-11: SNP profiles generated from the same sample but with different DNA concentrations; A) 3.3 ng/µl, and B) 0.3 ng/µl, electrophoresed with polymer POP-6™ using ABI 3500 PRISM® Genetic Analyzer…………………………………………………………………131

Figure 4-12: SNPs electropherogram of 52 markers generated from 1 ng 9947A…………..132

Figure 4-13: SNPs electropherogram of 52 markers generated from 1 ng tested individual sample………..133

Figure 5-1: (A) The SNPforID Browser homepage, and (B) several population databases that can be used for data comparison. [Taken from (SNPforID Consortium , Amigo et al. 2008a)……143

Figure 5-2: Several groups of populations that involved in SNP*for*ID project Taken from (SNPforID Consortium)…………………………………………………………………………………………146

Figure 5-3: Significant LD detected in the Malays…………………………………………………………157

Figure 5-4: Significant LD detected in the Chinese……………………………………………………..158

Figure 5-5: Significant LD detected in the Indians………………………………………………….….159

Figure 5-6: (A) Triangular plot classifying the individuals, (B) to (E) are density plot of the likelihood of all ethnic groups of data analyzed with the Snipper software. ……..…………..162

Figure 6-1: Electropherograms of Paternity Case 5 (mentioned in Table 6-1) involving putative father, mother and child: (A) SNaPshot profiles of 13_{th} multiplex assay of this trio, and (B) STR profile with one mutation occurrence at Child's *CSF1PO* marker…………..……178

Figure 6-2: STR profile of a Control family (Case 8- Table 6-1) involving father, mother and Child E, where two mutations had occurred on Child E at locus D8S1179 and FGA (as shown by the arrows)…………………………………………………………………………………………………179

Figure 6-3: Shown above are the parameters such as temperature, rainfall and relative humidity values in the month of August/September (for 27 days)………………………………180

Figure 6-4: Shown above are the simulated samples taken from Day 19 to Day 27. The arrows show the bloodstains on the cloths surrounded or covered by the fungi……………181

Figure 6-5: A) SNaPshot profiles of 13_{nd} multiplex assay of the simulated environmental samples from Day 1, Day 21 to Day 27, and (B) STR profiles of Powerplex® 16 system of these samples from Day 1, Day 21 to Day 27…………………………………………………………183

Figure 6-6: Number of profiles (STR and SNPs) that have been successfully developed or failed from the crime stains. For SNP, PP means sample developed more than 26 markers……186

Figure 6-7: Application of SNP and STR assays on the crime sample (with DNA concentration 2.95 ng/µl and IPC_{CT} 28). Full STR (16 markers) and SNaPshot (52 markers) profiles were generated from this sample…………………………………………………………………….187

Figure 6-8: Application of SNP and STR assays on the crime sample (with DNA concentration 15 ng/µl and IPC_{CT} undetermined). No STR loci was amplified but a full SNaPshot profile was generated from this sample (a portion of torn shirt bearing bloodstains)……………………………………………………………………………………………………….....188

LIST OF TABLES

Table 1-1: Commercially available multiplexes kits, with improved reagents [Taken from (Butler 2011)]. ………4

Table 1-2: Classifications of SNP markers [Taken from (Butler 2011)]……………………………17

Table 1-3: Classification of 2010 Census ethnic groups (Department of Statistics 2010, JAKOA 2013).]………27

Table 2-1: 52 SNP markers positions and coordinates (Sanchez et al. 2006a)…………………41

Table 2-2: PCR and SBE primers used in original study (Sanchez et al. 2006a)…………………42

Table 2-3: Several PCR reaction mix volumes used for the study to observe the sensitivity of the assay. ……………………………………………………………………………………………………….....48

Table 2-4: Combinations of primer concentrations in SNaPshot analysis………………………52

Table 2-5: New combination and arrangement of the same 52 SNP code markers (from the original study) into four sets of 13-plexes assays, *13st, 13nd, 13rd and 13th)*…………………..53

Table 2-6: PCR optimal conditions of single and multiplex reactions……………………………54

Table 2-7: Optimized final primers concentrations (μM) for each multiplex set………………55

Table 2-8: The optimal thermal cycling parameters……………………………………………….55

Table 2-9: Modified SBE length (bp)…………………………………………………………………57

Table 2-10: The optimal conditions used for the SBE reactions..58

Table 2-11: The final concentrations (µM) of SBE primers used in the SNaPshot reaction.59

Table 2-12: Shown below are the parameters for the analysis of PCR fragments……………61

Table 3-1: Table shows the combination of primers concentration used in the study……..89

Table 3-2: The SNP markers with primer concentrations (µM) used in the multiplex assays (*first trial*)……………………………………………………………………………………………………..91

Table 3-3: Multiplex assays consisting SNP markers with different SBE size and primer concentrations………………………………………………………………………………………………..93

Table 3-4: The first construction of four sets 13-plex PCR/SBE assays……………………………..95

Table 3-5: Second trial of 13-plex assays developed in the study…………………………………..97

Table 3-6: Four sets of 13-plexes assay reactions developed in this study……………………100

Table 3-7: A summary of alleles genotyped in 90 Malaysian individuals using TaqMan® SNP Genotyping assays which similar to the SNaPshot assay data……………………………………106

Table 4-1: The number of arrays in commercial CAE systems [Adapted from (Butler et al. 2004, Butler 2011)]………………………………………………………………………………………………114

Table 4-2: Polymer characteristics [Adapted from (Applied Biosystems 2012)….……………..117

Table 4-3: Run parameters for the SNP analysis on the 3500 Genetic Analyzer)……………117

Table 4-4: Data of allele's migration (bp) on different type of polymers………………………..123

Table 5-1: The populations included in the Sanchez study (Sanchez et al. 2006a)…………142

Table 5-2: Groups of populations used for comparison in this study from the (SNPforID Consortium)…………………………………………………………………………………………………..147

Table 5-3: Allele frequencies data of Malaysian and neighboring countries populations.149

Table 5-3: Allele frequencies data of Malaysian and neighboring countries populations (continued)……..150

Table 5-4: Observed (Ho) and expected [He] heterozygosities and P values from an exact test for HWE across 52 SNPs typed in 325 individuals from three ethnic groups……………152

Table 5-5: Comparison of Observed (Ho) and Expected (He) heterozygosities obtained in this study with the reference data (SNPforID Consortium)…………………………………………..153

Table 5-6: Average Expected heterozygosity (He) values of all 52 SNPs markers in different populations, p= 0.05…………………………………………………………………………………………...154

Table 5-7: Significant linkage disequilibrium at p<0.05 in Malays, Chinese and Indians…156

Table 5-8: Pairwise F-statistic (FST) values of 52 SNPs…………………………………………………160

Table 5-9: Percentages of population assignment using Snipper (A) and Arlequin version 3.5(B)…….161

Table 5-10: Use of 'cross verbose-validation' to compute the success ratio using the 52 SNPs with The Snipper software………………………………………………………………………………….163

Table 5-11: Forensic parameters calculated for 52 SNPs from each Malaysian ethnic group. RMP represents **random** match probability, PD represents power of discrimination, and PIC represents polymorphism information content……………………………………………………165

Table 5-12: Combined Paternity index and exclusion probability of 52 markers in Malays, Chinese and Indians group………………………………………………………………………………………166

Table 5-13: Forensic data comparison among the populations……………………………………167

Table 5-14: Forensic data collected and estimated for Malaysian populations……………. 168

Table 6-1: Paternity calculation using 16 STRs versus 52 SNPs markers in all kinship cases……..177

Table 6-2: Full alleles detected in Day 1, Day 21 to Day 27…………………………………………….182

Table 6-3: Results obtained from the analysis of 16 STRs and 52 SNPs on the casework samples. ……………………………………………………………………………………………………….185

Table 6-4: Statistical evaluation of SNPs data using DNA.VIEW software…………………… 189

ACKNOWLEDGEMENTS

There are many people I would like to express my very great appreciation to. I wish to thank them for their endless guidance and their enthusiastic encouragement which carried on throughout this whole project.

To begin with, I would like to offer my special thanks and gratitude to my supervisor Dr. William Goodwin and co-supervisor, Dr. Sibte Hadi for their valuable and their constructive suggestions during the planning and development of this entire PhD project. I would also like to thank Dr. Arati for her advice and assistance in especially statistical calculations. Correspondingly, I wish to say thank you to Dr. Judith, who also our Forensic group lab manager. Not to forget, my supportive RDT, Prof. Jai Paul and all staff from the Graduate Research School for their continuously guidance and help throughout my study and stay in Preston.

Secondly, my special thanks are extended to my colleagues; Sasitaran, Helen, Shahid, Chrissy, Shanti, Rashid, Haliza, Sheikha, Chris, Robin, Sophie, Michelle, Michael and others for keeping me updated with help, knowledge and support throughout my entire stay in Preston.

Last but not least, I am particularly grateful for the assistance given by my beloved family. They supported me greatly and stayed with me throughout my hardship to finished and cope with my study. Especially to my husband, Yusri Sabri and my children who have been nothing but encouraging and loving me always.

CHAPTER 1

INTRODUCTION

INTRODUCTION

1.1. GENETIC MARKERS IN THE FORENSIC DNA ANALYSIS

Since the introduction of DNA profiling in 1985, a broad spectrum of genetic markers and human identity-typing systems has emerged for forensic use. The core aim of forensic DNA testing is to generate a genetic signature that could either exclude or establish that there is an extremely high probability to infer a link between individual(s) and evidential material(s). Since the Colin Pitchfork murder investigation (Jeffreys et al. 1985), forensic science has successfully implemented a lot of molecular biology tools.

Today, a common set of standardized STR markers are widely used in forensic genetic casework. A large number of STR loci have been characterized but only around 20 are commonly analyzed in forensic casework (Goodwin et.al. 2011). Both Applied Biosystems and the Promega Corporation has become big providers of STR kits that permit robust multiplex amplification of the STR core markers (Butler 2011).

With the rapid growth of the human identification applications of STR loci, STRBase (http://www.cstl.nist.gov/boitech/strbase) was launched in 1997, to keep the forensic scientist up-to-date with the rapidly evolving field of DNA typing. It contains much useful information including population data for STR markers, links to other related web pages, new microvariants or “off-ladder” STR alleles and others (Butler 2011).

Various DNA technologies have since evolved, including lineage marker analyses which use maternally inherited mitochondrial DNA (mtDNA) hypervariable regions and paternal Y-chromosome polymorphisms (Y-STR). In forensic investigation, mtDNA analysis is very useful for generating profiles, particularly of human remains that are highly degraded and

not amenable to STR typing and also for kinship analysis, such tracing maternally –related families (Goodwin et.al. 2011, Coble et al. 2009). As for Y chromosomal markers, they are widely used for the forensic analysis of male DNA, especially in the analysis of mixed samples in rape cases (Jobling & King 2004). They are also important in paternity testing, historical investigations and exploring human migration patterns (Butler 2011). Another lineage marker that recently has gained the forensic scientist's attention is X-chromosome STR typing. This can be helpful in some kinship analysis situations, particularly in deficient paternity cases, where one of the parents is not available (Butler 2011).

There is another complementary technique to standard STR-based analysis which was also introduced and implemented in order to obtain DNA profiles from cells and debris transferred through skin contact, especially in cases such as snatch theft, housebreaking and car theft; this technique is called Low Copy Number (LCN) (Butler 2005). Low copy number (LCN) typing, particularly for current short tandem repeat (STR) typing, refers to the analysis of any sample that contains less than 200 pg of template DNA. Generally, LCN typing simply can be defined as the analysis of any DNA sample where the results are below the stochastic threshold for reliable interpretation (Budowle et. al. 2009).

However, there are several complex and specific challenges in generating DNA profiles such as there is no DNA profile or only partial DNA profile is amplified due to DNA degradation or the inhibitors of the biological samples. In order to address these problems several technologies including "miniSTRs" or reduced-sized PCR products were developed and introduced (Butler 2005). Same common STR loci used in commercially available kit exceed 400bp in size, but with "miniSTRs" the amplicons size is reduced, ranging from 50bp up to 280bp (Butler 2011).

In addition, to overcome PCR inhibitors and successfully generate DNA profiles, many more technologies and commercial kits with improved buffers were designed and developed by the providers as listed in Table 1-1.

Table 1-1: Commercially available multiplex kits, with improved reagents [Taken from (Butler 2011)].

Kit's Name	Source	Release date
AmpF1STR MiniFiler	Applied Biosystems	Mar 2007
AmpF1STR SEfiler plus (improved buffers)	Applied Biosystems	Nov 2007
PowerPlex 16 HS (same primers, improved reagents)	Promega	Mar 2009
PowerPlex ESX 16 & EXS 17	Promega	Sept 2009
PowerPlex ESI 16 & ESI 17	Promega	Sept 2009
AmpF1STR Identifiler Direct (same primers, improved reagents)	Applied Biosystems	Nov 2009
AmpF1STR Identifiler Plus (same primers, improved reagents)	Applied Biosystems	Jan 2010
AmpF1STR NGM	Applied Biosystems	Jan 2010
Investigator ESSplex	QIAGEN	April 2010
Investigator Decaplex SE	QIAGEN	April 2010
Investigator Triplex AFS QS	QIAGEN	April 2010
Investigator Triplex DSF	QIAGEN	April 2010
Investigator ESSplex	QIAGEN	Aug 2010
Investigator IDplex	QIAGEN	Sept 2010
Investigator HDplex	QIAGEN	Sept 2010
Investigator Hexaplex ESS	QIAGEN	Sept 2010
Investigator Nonaplex ESS	QIAGEN	Sept 2010
Investigator ESSplex SE	QIAGEN	Oct 2010
AmpF1STR NGM Select	Applied Biosystems	Dec 2010
PowerPlex 18D	Promega	Feb 2011

Although the core STR markers and other complementary markers have been extensively used in crime casework, yet the forensic DNA scientist often uses additional markers as the need arises to obtain more or further information in a particular sample or case (Butler 2011).

The study of single nucleotide polymorphisms (SNPs), has received a lot of attention in recent years, due to the abundance of SNPs throughout the human genome. The increasing interest in SNP typing in the forensic field, is not only for analyzing the geographical origin of samples, by defining Y chromosome or mtDNA haplogroups, but also for the potential applications of autosomal SNPs. This interest is growing particularly in the analysis of degraded DNA and complicated kinship testing (Sobrino et al. 2005a).

1.2. SINGLE NUCLEOTIDE POLYMORPHISMS (SNP)

A polymorphism has been defined as existing when at least two different alleles are present in a population, and both are present at ≥1% frequency. An allele below the 1% frequency used to define a polymorphism is sometimes called a *variant*, but this term is used by others to describe any form of DNA variation, irrespective of frequency or phenotypic effect. The literature also contains the frequency-independent term *single nucleotide variant (SNV)*, which some have argued should replace single nucleotide polymorphism (SNP). Our practice is to use “polymorphism” (including SNP) to indicate ≥1% allele frequency, “variant” for any less frequent difference between genome copies, and “mutation” to refer to a de novo change (Jobling et. al. 2013).

Approximately 7 billion people living today carrying about 14 billion genome copies, and, with the rare exceptions carried by identical twins, triplets, and so on, all of them are different. The differences between genomes encompass a wide range of scales: from single nucleotide polymorphisms (SNPs) through to structural variation involving millions

of base pairs of DNA sequence. Most differences have no discernible effects on the people who carry them, but some result in phenotypic diversity with no obvious deleterious effect, and some cause, predispose to, or protect against, serious diseases (Jobling et. al.2013).

The simplest and smallest-scale difference between two homologous DNA sequences is a *base substitution*, in which one base is exchanged for another (Figure 1-1 and Figure 1-2). When a pyrimidine base is exchanged for another pyrimidine (for example, C for T), or a purine for another purine (for example, A for G), the change is called a *transition*. When a pyrimidine is exchanged for a purine, or vice versa, this is a *transversion*. These differences are examples of single nucleotide polymorphisms, or SNPs (pronounced “snips”). The insertion or deletion (indel) of a single base is also included in the category of SNPs—perhaps unadvisedly, because the mechanisms which underlie these indels, and the ways they are analyzed, differ from those for base substitutions (Jobling et. al.2013).

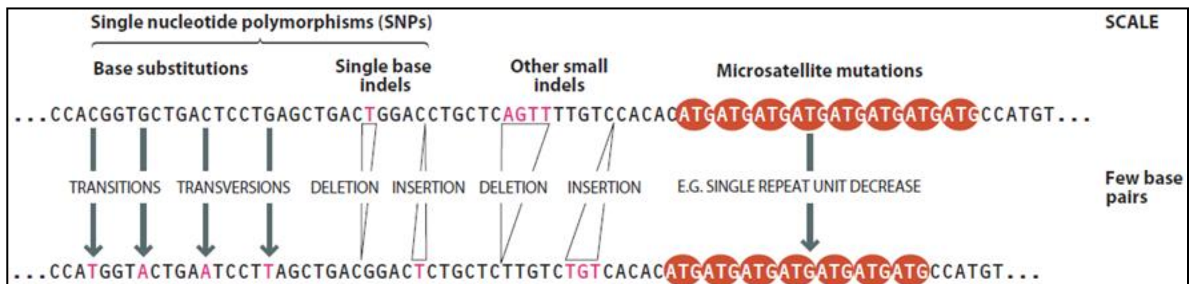


Figure 1-1: Examples of mutations affecting a few base pairs, including examples of transition and transversion substitutions [Taken from Jobling et.al. 2013].

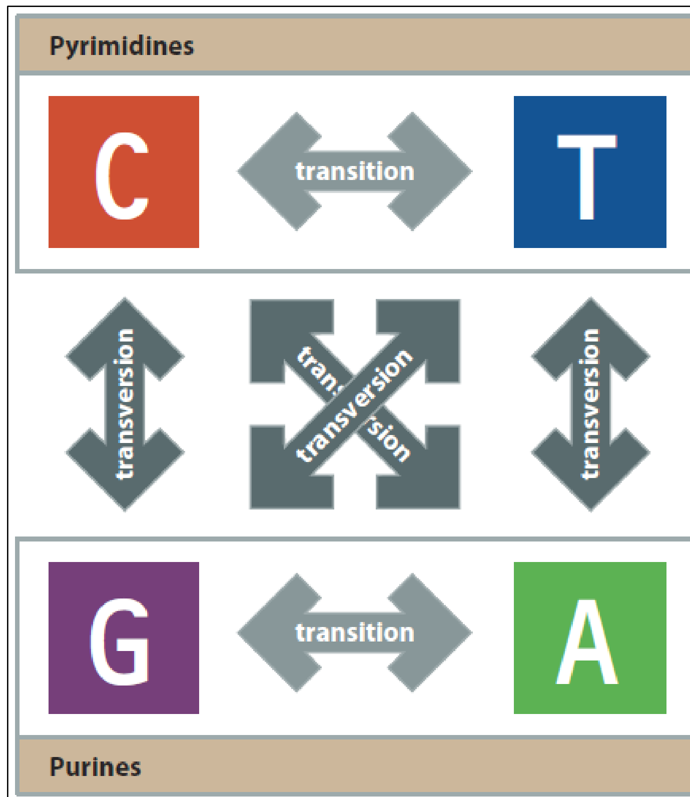


Figure 1-2: Transition and transversion mutations [Taken from Jobling et.al. 2013].

1.2.1. SNP Characteristics

SNPs have a number of characteristics that make them appropriate for forensic studies as well as population studies: (i) there are several millions of validated SNPs in the human genome; (ii) SNPs can be typed by semi-automated, cost-effective and standardized methods; (iii) Polymerase Chain Reaction (PCR) amplicons can be shorter than 100 bp, which allows for amplification and unambigous allele calling even from highly degraded DNA samples; (iv) they have very low mutation rates compared to short tandem repeats (STRs), a feature which is valuable for kinship testing; and, (v) as biallelic polymorphisms, SNPs are comparatively easy to validate. Although, individual SNPs have a discrimination power lower than that of individual STRs, it has been shown that 50 unlinked SNP loci with

high overall heterozygosity, matches or exceeds the discriminatory power of 15 STRs (Sanchez et al. 2006a).

Figure 1-3 shows mutation rates of micro- and minisatellites are very variable between loci. The rates given here reflect ascertainment bias, in that they are those of the mostly widely used and polymorphic loci. The average rates are probably much lower (Jobling et. al.2013).

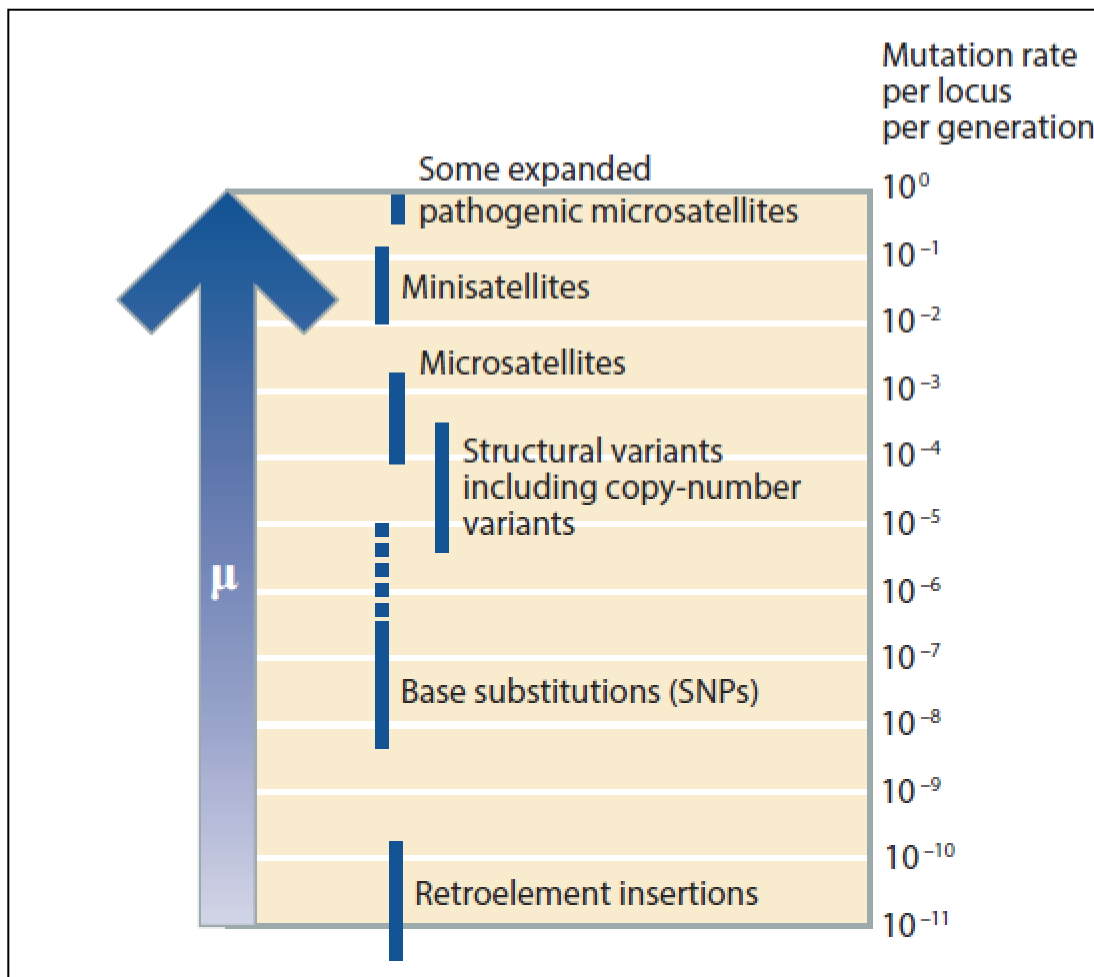


Figure 1-3: Overview of mutation rates (μ) of different classes of polymorphisms [Taken from Jobling et.al. 2013].

Base substitutions, occurring at an average rate of the order of 10^{-8} per base per generation, are about 10 times more frequent than indels, although this relative frequency varies substantially between loci, because it depends on the sequence context.

Transitions are almost three times more frequent than transversions, despite the fact that for a given base there are two possible transversions to only one transition (Figure 1-2)—if all mutations were equally probable we would expect a 2:1 ratio of transversions to transitions. This deviation from expectation might be explained by influences of efficiency of error detection and repair, sequence context, or perhaps by differences in misincorporation rates (Jobling et.al. 2013).

It is important to note that mutation rates of base substitutions are in general so low that a given mutation at a particular position is unlikely to have recurred or reverted in the small populations of early humans; independent occurrences are therefore not usually found today at appreciable frequencies. The low mutation rate also means that base substitutions generally belong to the class of binary polymorphisms, also known as biallelic, or diallelic polymorphisms, or unique event polymorphisms (UEPs) (Jobling et. al.2013).

Whilst, typical figures for mutation rate estimates from direct approaches are around 10^{-3} to 10^{-4} per polymorphic microsatellite per generation. Application of such rates to the dating of past events such as migrations and population expansions is somewhat controversial, since when populations with well-documented short-term histories are examined the effective mutation rate that explains the accumulated diversity seems to be considerably lower (Jobling et.al. 2013).

1.2.2. SNP Consortium/Databases

There are several groups and institutions that have contributed their efforts over the past few years in facilitating and characterizing SNPs, including the mapping of the entire human genome/ SNPs and also the development of SNP databases. Some of them are listed as below.

1.2.2.1. The European DNA Profiling Group (EDNAP)

EDNAP group (www.isfg.org/EDNAP) was started in 1988 by a group of forensic scientists from various parts of European with the main aim of harmonizing DNA technology for criminal investigations. This working group was accepted in 1991, at the Congress of the ISFG (International Society for Forensic Genetics). Since then, this group has been working together and collaborates in many forensic DNA research projects and analyses.

1.2.2.2. The SNP Consortium Ltd. (TSC)

TSC was established in 1999, a collaboration between major pharmaceutical companies, the Wellcome Trust, and five leading academic and genome sequencing centers (*http://snp.cshl.org*). TSC has produced several million mapped and characterized human SNP markers, in samples from African, European and Asian populations (Thorisson & Stein 2003, Butler 2005). TSC is one of the biggest contributors to the HapMap Project.

1.2.2.3. HapMap Project

The International HapMap Project, launched in 2002, is a partnership of scientists and funding agencies from Canada, China, Japan, Nigeria, the United Kingdom and the United States to identify and catalog genetic similarities and differences in human beings.

The aim of the International HapMap Project was to characterize patterns of genetic variation and LD in a sample of 269 individuals from four geographically defined populations with ancestry in Africa, Europe, and Asia, thus facilitating the design of efficient genome wide association studies for common disease. The study began in 2002, and by 2007 had published a haplotype map based on over 3.1 million SNPs (Jobling et. al.2013).

The HapMap Web site provides researchers with a number of tools that allow them to analyze the data as well as download data for local analyses. In phase I of the project, more than 1.1 million SNPs were genotyped in 269 individuals from four worldwide populations (Yoruba, Japanese, Chinese and European) (Thorisson et al. 2005).

In Phase II, over 3.1 million human SNPs were characterized and genotyped from those 269 individuals. Data shows that 10–30% of pairs of individuals within a population share at least one region of extended genetic identity arising from recent ancestry and that up to 1% of all common variants are untaggable, primarily because they lie within recombination hotspots. The recombination rates found varied systematically around genes and between genes of different function (Frazer et al. 2007).

The International HapMap Project 3 together with the Human Genome Project and the SNP Consortium has also successfully identified 10 million common DNA variants, primarily SNPs, in a limited set of DNA samples. Knowledge of these SNPs and their linkage-disequilibrium (LD) patterns enabled genome-wide association studies, which have successfully identified hundreds of novel genomic loci that influence human diseases (Altshuler et al. 2010). Linkage disequilibrium (LD) is the association of alleles because of

reduced recombination between them, which can reflect either physical distance, or variable recombination activity (Jobling et. al.2013).

The results of HapMap demonstrate an extremely non-uniform distribution of recombination in the genome, reflected in a discontinuous block-like structure of LD. These haplotype blocks can contain many SNPs that are highly correlated with one another, and display the limited haplotype diversity that is expected from relatively few historical recombination events. The breakdown of LD between blocks is often abrupt, reflecting the occurrence of most human recombination events within hotspots. Computational methods allow the local recombination rate to be calculated from the population-scale genotypic data across the whole genome: rates vary greatly, and from HapMap Phase II data approximately 33,000 hotspots showing greatly elevated recombination could be identified, with a mean width of 5.5 kb (constrained by SNP density— most empirically determined hotspots are considerably smaller). About 60% of recombination occurred within these hotspots, yet they comprised only ~6% of sequence (Jobling et.al.2013).

In phase III, an extended set of 1,184 samples from 11 populations were studied which 1.6 million common SNPs were genotyped. The additional populations were from African ancestry in the southwestern USA (ASW), Chinese in metropolitan Denver, Colorado, USA (CHD), Gujarati Indians in Houston, Texas, USA (GIH), Luhya in Webuye, Kenya (LWK), Maasai in Kinyawa, Kenya (MKK), Mexican ancestry in Los Angeles, California, USA (MXL), and Tuscans in Italy (Toscani in Italia, TSI) (Altshuler et al. 2010).

1.2.2.4. HGDP-CEPH Diversity Panel Database

This panel database (V3.0) is designed to receive and store polymorphic marker genotypes, copy number variant (CNVs) calls, and Sanger DNA sequences generated by users of the DNAs of the HGDP-CEPH Diversity Panel. There are 11 data sets in this panel and the data are publically accessible via this website (http://www.cephb.fr/en/hgdp/).

This panel was initiated by a group of scientists at Stanford University that collaborated with the Center for the Study of Human Polymorphism (CEPH) at the Fondation Jean Dausset in, Paris. They have studied a large scale of genetic diversity in human populations. Genomic DNA was analyzed from more than 1,000 individuals from 52 different populations, determining their genotypes at more than 650,000 SNP loci, with the Illumina BeadStation technology. The collection of these DNA is also referred to as the "HGDP-CEPH Human Genome Diversity Cell Line Panel". The indigenous populations involved are from Africa, Europe, the Middle East, South and Central Asia, East Asia, Oceania and America (Cann et al. 2002, Rosenberg et al. 2002, Cavalli-Sforza 2005, Li et al. 2008).

Many investigators and researchers have been using these cell lines to carry out their studies. In order to establish a 34-plex ancestry-informative marker (AIM) assay, a group of scientists (Phillips et al. 2007b) have used three main population-group HGDP-CEPH samples of geographically-known origin to estimate the misclassification rates by statistical means in predicting ancestral origin of the training sets populations (60 Mozambican, 60 Somali, 60 Galician, 60 Danish, 60 Mainland Chinese and 60 Taiwanese). In addition, several reference population of HGDP-CEPH were also used in the revision of an established 34 SNP forensic ancestry and for comparison in analyzing four U.S. populations which show a range of admixture patterns (Fondevila et al. 2013a).

1.2.2.5. 1000 Genomes

The 1000 Genomes Project is an international collaboration to produce an extensive public catalog of human genetic variation, including SNPs and structural variants, and their haplotype contexts. The goal of the 1000 Genomes Project is to find most genetic variants that have frequencies of at least 1% in the populations studied and to estimate the frequencies of the variant alleles, and also identify the haplotype and characterize their LD patterns.

The 1000 Genomes Project Consortium targeted about 2500 anonymous people from about 25 populations around the world which have been sequenced using next-generation sequencing technologies.

In Phase 1, the genomes of 1,092 individuals from 14 populations were constructed using a combination of low-coverage whole-genome and exome sequencing. The Consortium discovered and genotyped 38 million single nucleotide polymorphisms, 1.4 million short insertions and deletions, and more than 14,000 larger deletions.

The project's pilot phase used population samples collected for the HapMap project and comprised three sub-projects as follows:

• *Trio pilot*: whole-genome sequencing at high coverage (each base covered on average 42 times) of two families of African and European origins (YRI and CEU), each including two parents and one daughter.

• *Low-coverage pilot*: whole-genome sequencing at low coverage (2–6×) of 59 unrelated African individuals (YRI), 60 unrelated European individuals (CEU), and 60 unrelated East Asian individuals (30 CHB and 30 JPT).

• *Exon pilot*: targeted high-coverage (average >50×) sequencing of 8140 exons from 906 randomly selected genes (total of 1.4 Mb), in 697 individuals from seven populations of African (YRI, LWK), European (CEU, TSI), and East Asian (CHB, JPT, CHD) ancestry.

The data generated by the 1000 Genomes Project not only aid the interpretation of all genetic-association studies, but also provide lessons on how best to design and analyze sequencing-based studies of disease. A further 1,500 individuals from 12 new populations, including at least 15 high-depth trios; will form the final phase of this project (Altshuler et al. 2012).

1.2.2.6. The Single Nucleotide Polymorphism Database (dbSNPS)

The dbSNPs or dbSGV (database of short genetic variations) is a public- domain archive for a broad collection of simple genetic polymorphisms and it was developed to complement GenBank. This collection of polymorphisms includes single-base nucleotide substitutions (also known as SNPs), multi-base deletions or insertions, retro-transposable element insertions and microsatellite repeat variations (STRs).

The SNP database can be queried from the dbSNP homepage by using Entrez SNP, or by using the links to the six basic dbSNP search options stated in the NCBI handout (ftp://ftp.ncbi.nlm.nih.gov/pub/factsheets/Factsheet_SNP.pdf). By 2008, there were about 12.8 million human reference SNPs have been deposited into the public database dbSNP build 128 (http://www.ncbi.nlm.nih.gov/books/NBK44423/).

1.2.2.7. The SNPforID Browser (SNPforID Consortium)

Several members of the European forensic DNA typing community launched a project in 2003 known as SNP*for*ID that is developing SNP assays to directly aid forensic DNA analysis (Phillips et al. 2004). This group is endeavouring to develop highly multiplexed SNP assays using unlinked loci that are well spread throughout the human genome for high-throughput forensic identification of individuals. Population data are also being gathered to measure SNP allele frequencies in various groups (Butler 2005). The SNP*for*ID browser page contains a world map which is clickable at any population selected. There are two main sets of SNPs available: SNP*for*ID 52-plex (for individual identification markers) and SNP*for*ID 34-plex (ancestry markers). The website page also contains the frequencies tab, statistic tabs and download tabs, where users are allowed to retrieve the alelle frequencies and other forensic parameters such as hetereozygosity values, F_{ST} etc. The SNP*for*ID studies were carried out on populations distributed across Africa, America, Europe, East-Asia, Central-South Asia, Middle East and Oceania.

In order to facilitate the SNPforID Browser user with multiple databases, which allow the user to compare or access them easily, a novel web tool has been developed and termed SPSmart (SNPs for Population Studies). These databases include HapMap, Perlegen, CEPH (Michigan and Stanford) and 1000 Genomes (Amigo et al. 2008b).

1.2.3. Categories of SNP Markers

At the 22nd Congress of the International Society of Forensic Genetics (ISFG) which was held in Copenhagen, Denmark on 2007, there was a panel discussion on single nucleotide polymorphisms (SNPs) and their application in forensic identity and relationship testing. During the meeting, the scientists' experiences in SNPs were presented on their developed SNP panels, and they discussed the benefits, limitations and requirements for SNP implementation and interpretation (Butler et al. 2008, Budowle, van Daal 2008, Butler 2011). As a result, the scientists and the panel of experts have decided to classify the SNP markers into four general categories, as shown in Table 1-2.

Table 1-2: Classifications of SNP markers [Taken from (Butler 2011)].

Category	Characteristics	Examples
Identity SNPs: Individual identification SNPs (IISNPs)	SNPs that collectively give very low probabilities of two individuals having the same multi-locus genotype.	FSS 21plex (Dixon et al. 2005)
		SNPforID 52plex(Sanchez et al. 2006a)
		Kidd group SNPs(Pakstis et al. 2010)
Lineage SNPs: Lineage Informative SNPs (LISNPs)	Sets of tightly linked SNPs that function as multi-allelic markers that can serve to identify relatives with higher probabilities than simple bi-allelic SNPs.	mtDNA coding region SNPs (Coble et al. 2004)
		Japanese Y-SNPs (Mizuno et al. 2010)
		Haplotype blocks (Ge et al. 2010)
Ancestry SNPs: Ancestry Informative SNPs (AISNPs)	SNPs that collectively give a high probability of an individual's ancestry being from one part of the world or being derived from two or more areas of the world	SNPforID 34-plex (Phillips et al. 2007b)
		24 SNPs (Lao et al. 2010)
		FSS YSNPs (Wetton et al. 2005)
Phenotype SNPs: Phenotype Informative SNPs (PISNPs)	SNPs that provide a high probability that the individual has particular phenotypes, such as a particular skin colour, hair colour, eye colour, etc.	Red Hair (Grimes et al. 2001)
		"Golden" gene pigmentation (Lamason et al. 2005)
		IrisPlex eye colour (Walsh et al. 2011)

1.2.4. Widely-used SNP Typing Methodologies

A number of SNP typing methods are available, each with their own strengths and weaknesses (Sobrino et al. 2005a). But among those methods, few are widely-used by the researchers or forensic scientist, this could be possibly because most of the forensic laboratories are already well-equipped with particular high-throughput instruments, such as capillary electrophoresis (CE) and Real-time PCR.

1.2.4.1. Primer extension (PE)

PE is one of the most frequently used detection methods for SNP genotyping and is also known as minisequencing and single base primer extension. The mechanism is based on the ability of DNA polymerase to incorporate specific deoxyribonucleotides complementary to the sequence of the template DNA. The product can be detected by capillary electrophoresis (CE) analysis as described in the SNaPshot® method (Applied Biosystems 2009a). By using this method, multiplexing of SNPs marker is feasible and practicable but a lot of work is needed to optimize the PCR and SNaPshot reaction of each set of SNPs. Several studies on autosomal SNPs, Y-SNPs, mtDNA-SNPs and ancestry SNPS (Sanchez et al. 2003, Sanchez et al. 2006a, Brión et al. 2005, Tomas et al. 2010, Gusmao et al. 2012, Fondevila et al. 2013a, Ballantyne et al. 2012) have been undertaken in different forensic laboratories with the analysis of SNPs using this technology (Sobrino et al. 2005a).

1.2.4.2. Oligonucleotide Ligation Assay (OLA)

OLA is a SNP typing method based on the ability of ligase to covalently join two oligonucleotides when they hybridize next to one another on a DNA template. The OLA assay requires that three probes are designed, one common and two allele specific. The

common probe anneals to the target DNA immediately downstream of the SNP. One allelic probe has at the 3′ end, the nucleotide complementary to one allele, with the other allelic probe having the nucleotide complementary to the alternative allele. These two allelic probes compete to anneal to the DNA target adjacent to the common probe. This generates a double-stranded region containing a nick at the allele site. Only the allelic probe perfectly matched to the target will be ligated to the common probe by DNA ligase. The multiplex capability of several technologies based on OLA is very high, but it is not suitable for forensic casework because it requires a high amount of DNA template (37ng) (Sobrino et al. 2005a). Several SNPs studies have been carried out with OLA technology such as the SNPlex™ Genotyping system (De La Vega et al. 2005a) and Genplex™ Typing system (Phillips et al. 2007a, Tomas et al. 2008, Tomas et al. 2011, Kidd et al. 2012).

1.2.4.3. TaqMan® SNP Genotyping Assays

The TaqMan® SNP Genotyping Assay is a single-tube PCR assay that exploits the 5′ exonuclease activity of AmpliTaq Gold® DNA Polymerase. The assays includes two locus-specific PCR primers that flank the SNP of interest, and two allele-specific oligonucleotide TaqMan® probes. These probes have a fluorescent reporter dye at the 5′ end, and a non-fluorescent quencher (NFQ) with a minor groove binder (MGB) at the 3′ end. The use of two probes, one specific to each allele of the SNP and labeled with two fluorophores, allows detection of both alleles in a single tube. Fluorogenic probes with an MGB produce enhanced allelic discrimination, because the MGB stabilizes the double-stranded probe template structure, thereby increasing the probe melting temperature (Tm) without increasing probe length. This provides enhanced mismatch discrimination between these shorter probes, resulting in improved specificity. All MGB probes include a non-fluorescent quencher (NFQ) that virtually eliminates the background fluorescence associated with traditional quenchers (De La Vega et al. 2005a). A few studies (Kidd et al. 2006, Pakstis et al. 2007, Li et al. 2010) have used this approach in screening their SNP

marker candidates and others (Phillips et al. 2007a, Harayama et al. 2011, Dulik et al. 2012) used TaqMan technology as for comparison in data obtained.

1.3. SNPS AS FORENSIC MARKERS

Often, forensic samples will contain too little template DNA or too degraded DNA and require alternate genetic marker analyses or approaches to what is currently used for routine casework. Single nucleotide polymorphisms (SNPs) have promising features to support forensic DNA analyses, such as the abundance of potential markers, amenability to automation, and potential reduction in required fragment length to only 60-80 bp (Budowle, van Daal 2008).

Budowle and van Daal (2008) also suggested that the SNP markers will serve an important role in analyzing challenging forensic samples, such as those that are greatly degraded, for augmenting the power of kinship analyses and family reconstructions for missing persons and unidentified human remains, as well as for providing investigative lead value in some cases without a suspect (and no genetic profile match in CODIS).

1.3.1. Degraded DNA

The primary point of SNPs in the forensic field is the ability to overcome challenges of strong inhibitors or samples possessing highly degraded material. PCR products for SNP markers can be small because the target region is a single nucleotide and amplicons are usually less than 100 bp in length. For these reasons, SNPs are likely to contribute more to individual identification especially for mass disaster victim identification. For example, SNPs have successfully been used on highly degraded samples from victims of the World Trade Center disaster of 11 September 2001 (Butler 2005).

1.3.2. Human Identification

Most SNPs are bi-allelic and thus are not as informative on a per-locus basis as the forensically selected STR loci. Because all forensic DNA indices, particularly convicted felon DNA databases, are well-established and based on STR loci, it is unlikely that SNPs will become the primary forensic markers for human forensic identification (Budowle, van Daal 2008, Gill et al. 2004).

However, SNPs do have advantages for forensic analyses in some situations, such as use in mass disasters and missing person's cases where the DNA may be substantially fragmented. Lastly, SNPs have relatively low mutation rates which are useful for lineage-based analyses, such as inheritance cases, missing person cases, and situations where no direct reference sample may be available (Budowle, van Daal 2008).

In order to obtain the same power of discrimination as existing STR multiplex systems, a panel of at least 50–100 autosomal SNP loci would be required (Chakraborty et al. 1999, Gill 2001). However, creating such large multiplexes that are validated for forensic analysis is a challenging task (Budowle, van Daal 2008).

1.3.2.1. SNP panels of 40-100 plex assays for human identification

Millions of SNPs have been characterized in the human genome; however not all of these are suitable for forensic analysis. Based on selection criteria such as high levels of observed heterozygosity (average >0.45) and low F_{ST} (average <0.06) between populations several sets of markers have been developed for human identification. A number of SNP platforms have been developed to be applied for the detection of these SNPs markers.

1.3.2.1.1. SNPforID 52plex markers

A multiplex of 52 SNP markers has been developed and these markers have been reported to be polymorphic in European, Asian and African populations. The amplification of all 52 DNA fragments was carried out in one PCR reaction followed by detection of the SNPs with two single-base extension (SBE) reactions, which later detected by Capillary Electrophoresis (CE). In this study, these SNPs were also efficiently used and successfully amplified the degraded samples, where previously only partial STRs profiles had been obtained (Sanchez et al. 2006a, Sanchez et al. 2006b).

1.3.2.1.2. SNPlex™ Genotyping System

The SNPlex™ Genotyping system (Applied Biosystems) has been developed for applications that require scalable, high-throughput SNP genotyping capabilities. It utilizes a suite of pre-optimized universal assay reagent kits and a set of SNP-specific ligation probes to perform genotyping up to a 48-plex level in a single reaction. This system is based on the oligonucleotide ligation/ PCR assay (OLA/PCR) with a universal ZipChute™ probe detection for high-throughput SNP genotyping (De La Vega et al. 2005a). However, this system is no longer available commercially.

1.3.2.1.3. GenPlexTM Typing System

The Genplex™ Typing system 49plex assay has been developed based on the modification of SNPlex™ chemistry (both Applied Biosystems; AB SNPlex™ chemistry guide) using oligo-ligation of pre-amplified DNA and dye-labelled, mobility-modified detection probes. This gives highly predictable electrophoretic mobility of the allelic products generated from the assay to allow detection with standard CE analyzers. Few panels of SNPs markers have been designed using the GenPlex typing systems, such as the 48 most informative autosomal SNPs from the SNP*for*ID core discrimination set supplemented with the

amelogenin sex marker. These SNPs are evenly distributed across all 22 autosomes, exhibit balanced polymorphisms in three major population groups (African, Asian and European) and have been previously shown to be effective markers for forensic analysis (Phillips et al. 2007a). Another two panels of 49 and 41 SNPs were also used to test this system (Pakstis et al. 2010, Kidd et al. 2012). However, this system has been discontinued and is not available commercially.

1.3.2.1.4. Kidd Group IISNPs

Initially, Kidd et. al (2006) proposed a preliminary panel of 19 SNPs for the identification of humans based on their evaluation of match probabilities. They tested these 19 SNPs on a broad representation of world populations. This 19-SNP panel is commercially available through Applied Biosystems and has been proven to be high throughput, highly accurate, time-efficient and cost-effective (Kidd et al. 2006, Applied Biosystems 2010a). Later, Kidd and colleagues studied 40 SNP markers that were suggested to be suitable candidates for the global forensic community for a universally applicable SNP panel for human identification (Pakstis et al. 2007). Kidd and colleagues again extended their panel from 40 to 92-IISNPs based on the 44 populations studied. However, the final full set of 86-IISNPs was chosen to be used because no significant Linkage Disequilibrium (LD) was found across the 44 populations studied (Pakstis et al. 2010, Kidd et al. 2012).

1.3.3. Estimating ethnic origin of a sample

With the advent of dense maps of human genetic variation, an increasing amount of information is being assembled, which together with new analytical methods, is making it possible to explore the recent evolutionary history of the human population (Sabeti et al. 2007). In many cases, the identification of the population group from which a crime scene sample has come from can be valuable intelligence for investigating agencies. Panels

consisting of mtDNA SNPs and Y SNPs have already been found useful for this purpose but are intrinsically limited by the fact that they can only provide information on either the maternal or paternal ancestry. Autosomal SNPs that have different frequencies in different major population groups can provide valuable information on geographic ancestry (Goodwin et.al. 2011).

1.4. APPLICATION OF SNPs MARKERS ON THE MALAYSIAN POPULATIONS AND FORENSIC CASEWORK

1.4.1. Geography and demographic of Malaysia

The Federation of Malaya, consisting of 11 states in what is now Peninsular Malaysia, became independent from British rule in 1957. Sabah, Sarawak and Singapore joined the federation to form Malaysia in 1963; Singapore then left the federation to become an independent nation two years later. Since then Malaysia comprises Peninsular Malaysia, Sabah and Sarawak (Figure 1-4).

Migration is one of the defining issues of the twenty-first century in Malaysia. The multi-racial population in Malaysia is evidence of earlier inward migration during the nineteenth and the early twentieth century. Colonial rule saw the entry of foreign labour especially from China and India to work in the plantations, tin mining and the construction sectors.

Population distribution varies by state: Selangor is the most populous state (5.46 million), followed by Johor (3.35 million) and Sabah (3.21 million). The population share of these states to the total population of Malaysia was 42.4 per cent. The least populated states were federal territories, Putrajaya (72,413) and Labuan (86,908).

Population density of Malaysia stood at 86 persons per square kilometer in 2010 compared with 71 persons in 2000. Unlike the population distribution, the population density revealed a different picture. Selangor, being the most populous state, was only ranked fifth in terms of population density with 674 persons per square kilometer. Among the most densely populated states were Kuala Lumpur (6,891 persons), Pulau Pinang (1,490 persons) and Putrajaya (1,478 persons).

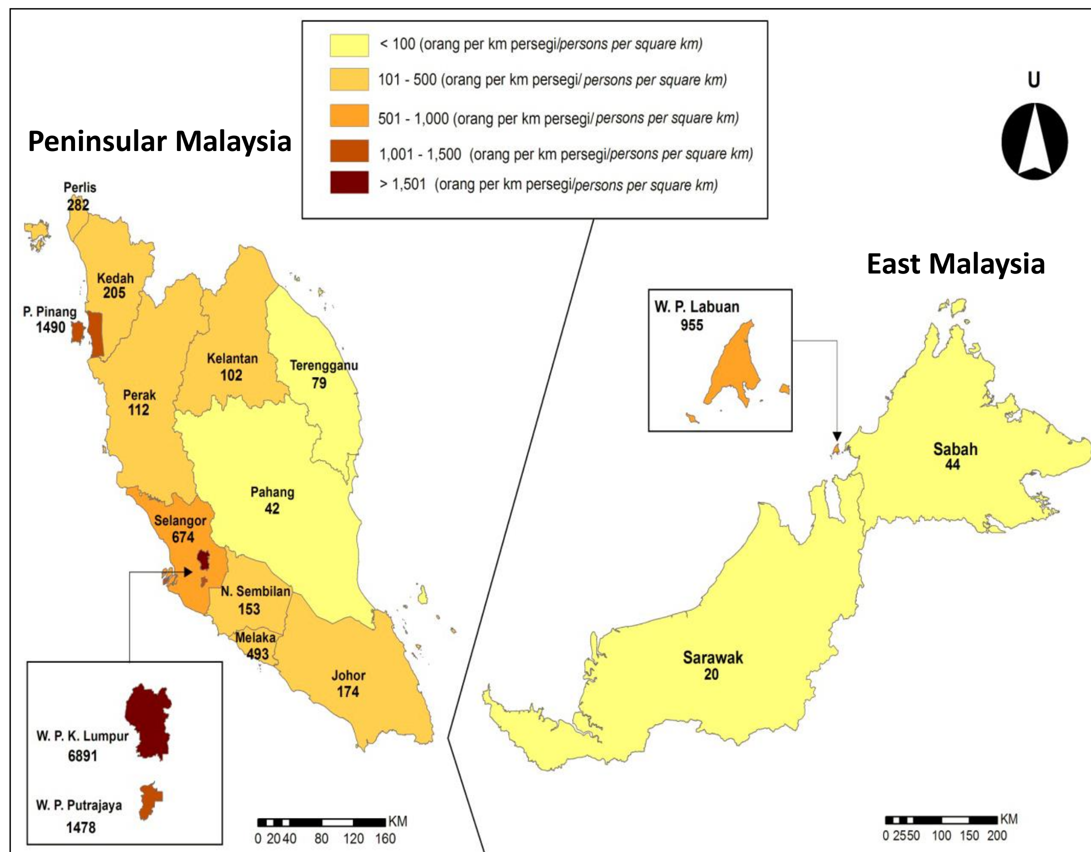


Figure 1-4: Population density by state Malaysia, 2010 [Taken from Department of Statistics 2010].

1.4.2. Multi-racial residents in Malaysia

Malaysia is a multi-racial country with more than 70 identified ethnic groups. These ethnic groups are broadly classified into four major groups namely Bumiputera (inclusive of Malay and Indigenous), Chinese, Indians and Others (including non-Malaysian citizen). The classification of the registered ethnic groups is listed in Table 1-3.

The population of Malaysia in 2010 was 28.3 million with an annual population growth rate of 1.9 per cent. The population is estimated to reach 31.6 million in 2020 (Zarinah Mahari 2011). Bumiputera, the main ethnic group constituted 67% per cent of the total population followed by Chinese at 24%, Indians 7% and others 0.7% (Department of Statistics 2010).

1.4.3. Forensic Casework in Malaysia

The Forensic Division of Kimia Malaysia is a national leader in the forensic science community that provides independent and impartial forensic science services to both government agencies and to the private sector. DNA profiling analysis is conducted only in the Headquarters in Petaling Jaya, Kuching Branch Laboratory, and Penang Branch Laboratory (www.kimia.gov.my). In addition, the Division also provides other scientific support including expert opinions, giving lectures and training and is involved in the collaboration research with the universities. The Forensic Division has been accredited by the ASCLD/LAB® (American Society of Crime Laboratory Directors/ Laboratory Accreditation Board) to the international quality standard ISO 17025 since 2005.

Table 1-3: Classification of 2010 Census ethnic groups (Department of Statistics 2010, JAKOA 2013).

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

1.5. STUDIES RELATED TO MALAYSIA

1.5.1. Autosomal SNP studies on Southeast Asian and East Asian populations

The HUGO Pan-Asian SNP Consortium has reported a large-scale study on the autosomal variation from various geographic samples of Asian populations. They found that genetic ancestry had strong correlation with linguistic associations as well as geography. These genome-wide studies of SEA and EA populations' were carried out using the Affymetrix GeneChip HumanMapping 50K Xba Array. Stringently quality-controlled genotypes were obtained at 54,794 autosomal single-nucleotide polymorphisms (SNPs) in 1928 individuals representing 73 Asian and two non- Asian HapMap populations (The International HapMap Consortium 2003).The study showed that about 50% of East Asian (EA) haplotypes were found in Southeast Asian (SEA) with only 5% in Central-South Asian (CSA) populations. This indicates that SEA is a major geographic source of EA populations.

The phylogenetic tree shown in Figure 1-5(A), demonstrates that SEA and EA populations share a common origin. These observations suggest that the geographic source(s) contributing to EA populations were mainly from SEA populations; with rather minor contributions from CSA, whilst the CSA populations showed they are sharing ancestry with European populations.

From the maximum likelihood phylogenetic tree (Figure 1-5(B)), the Negritos of the Philippines (PI-IR, PI-MW, PI-AT, PI-AG and PI-AE) and Malaysia (MY-KS and MY-JH) differ from neighbouring populations. The results suggested that there is substantial genetic proximity of SEA and EA populations (Abdulla et al. 2009). A very similar pattern of the population tree has been reported by Li et al. (2008) based on their previous study with 642,690 SNPs (Li et al. 2008) .

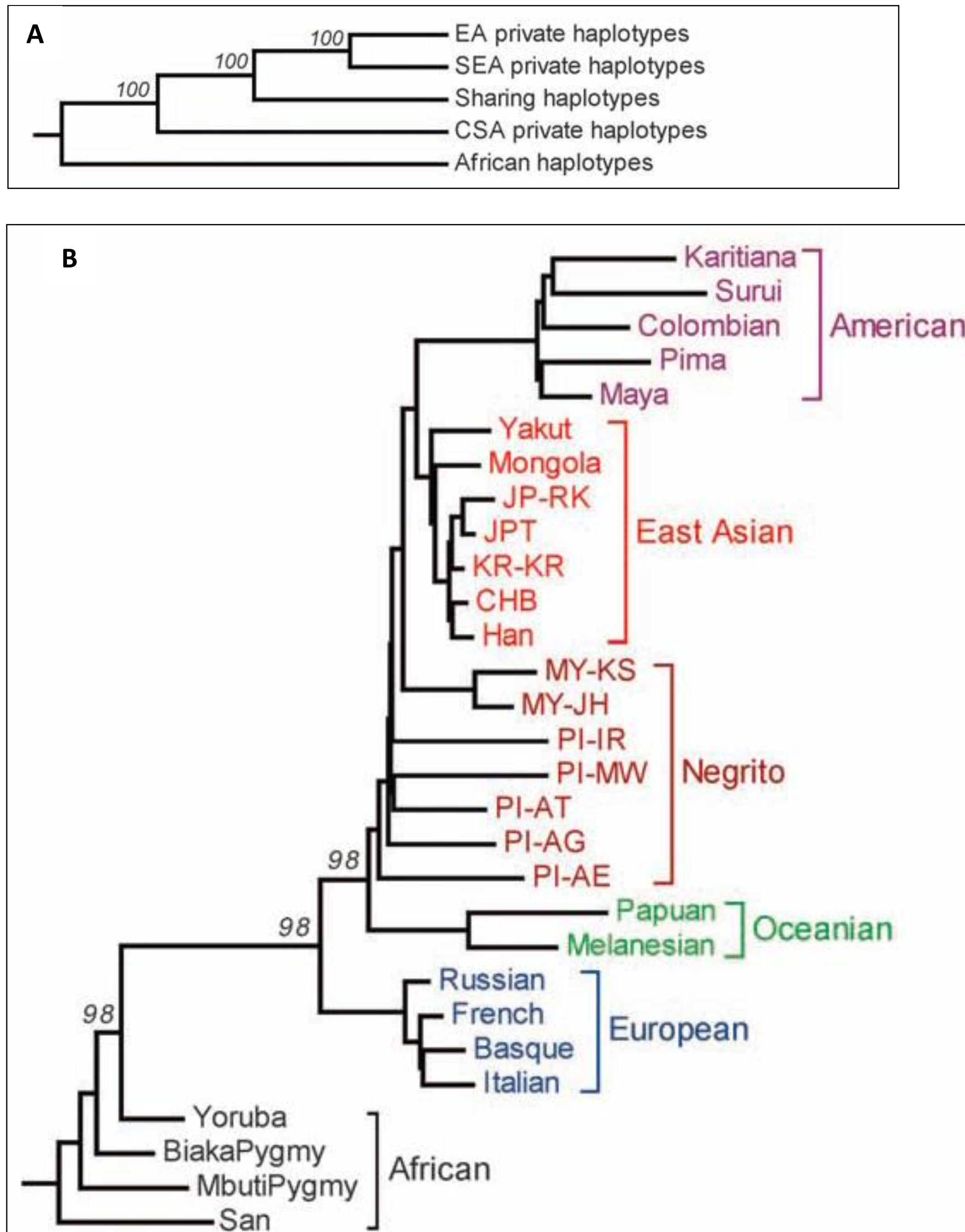


Figure 1-5: Numbers on the tree branches are bootstrap values. (A) Phylogeny of group private haplotypes. EA and SEA private haplotypes demonstrates sharing a common origin. Shared haplotypes are found in all EA, SEA, and CSA samples; African haplotypes were only used as out-group. (B) Maximum likelihood tree of 29 populations. The tree is based on data from 19,934 SNPs. [Taken from (Abdulla et al. 2009)].

Other recent studies have investigated the genetic structure of the Malay population in Malaysia using 54,794 genome-wide SNPs genotype data generated from four sub-ethnic groups: Peninsular Malaysia (Melayu Kelantan (MY-KN), Melayu Minang (MY-MN), Melayu Jawa (MY-JV) and Melayu Bugis (MY-BG)) (Hatin et al. 2011). These data were compared to 11 other populations from Indonesia, China, India, Africa and indigenous populations in Malaysia, which were obtained from the Pan-Asian SNP database. The results from the study suggest the Malay sub-ethnic groups are separated into at least three different clusters as shown in Figure 1-6. The MY-JV, MY-BG and MY-MN have a very close genetic relationship with Indonesian populations (ID-TR, ID-ML and ID-JV) indicating a common ancestral history, while the MY-KN formed a distinct group on the tree indicating that they are genetically divergent from the other Malay sub-ethnic groups. In Cluster II, MY-JV was also found to be closely related to the Proto-Malays (MY-TM) and Chinese (CN-JN and CN-WA). Both Chinese populations (CN-JN and CN-WA) were originated from Yunnan, China.

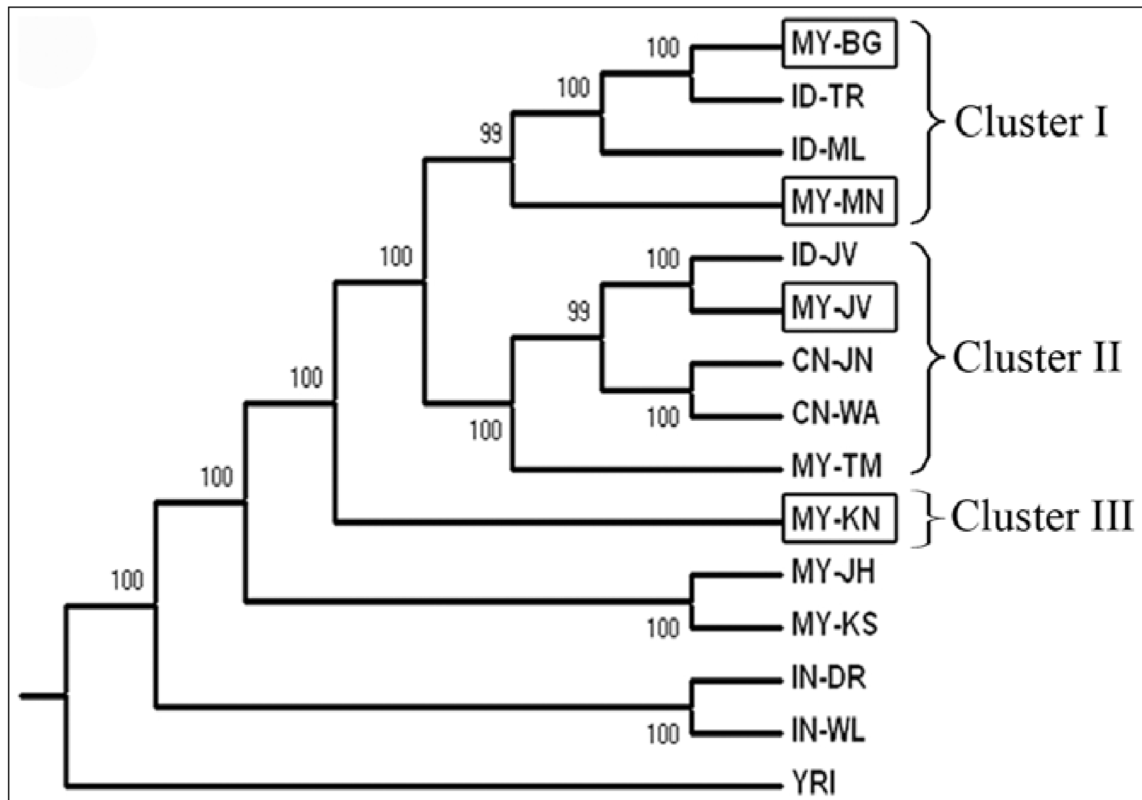


Figure 1-6: Neighbour-joining tree on 15 populations based on the F_{ST} measurement. Malaysia populations: Melayu Kelantan (MY-KN), Melayu Minang (MY-MN), Melayu Jawa (MY-JV), Melayu Bugis (MY-BG), Proto-Malay (MY-TM), Jahai Negritos (MY-JH) and Kensui Negritos (MY-KS); Indonesia populations: Indonesian Toraja (ID-TR), Indonesian Melayu (ID-ML) and Indonesian Jawa (ID-JV); China populations: Chinese Jinuo (CN-JN) and Chinese Wa (CN-WA); India populations: Marathi (IN-WL) and Telugu (IN-DR); African: Yoruba (YRI) [Taken from (Hatin et al. 2011)].

1.5.2. Theories on migration routes within Asia

Scientists have produced two basic theories to explain the initial peopling of Asia based on the research carried out by the HUGO Pan-Asian SNP Consortium. As shown in Figure 1-7, there are thought to have been two major waves of migration from the Middle East: one wave followed a southern coastal route, around the border of India, and continued from island to island across Malaysia, Indonesia, and the Philippines to the Pacific (see dark green arrows); a second wave is thought to have occurred eastwards across the Eurasian steppe turning south through the Asian mainland. A second theory suggests just one initial migration: along the coastal route, with populations moving north into East Asia from there (Normile 2009).

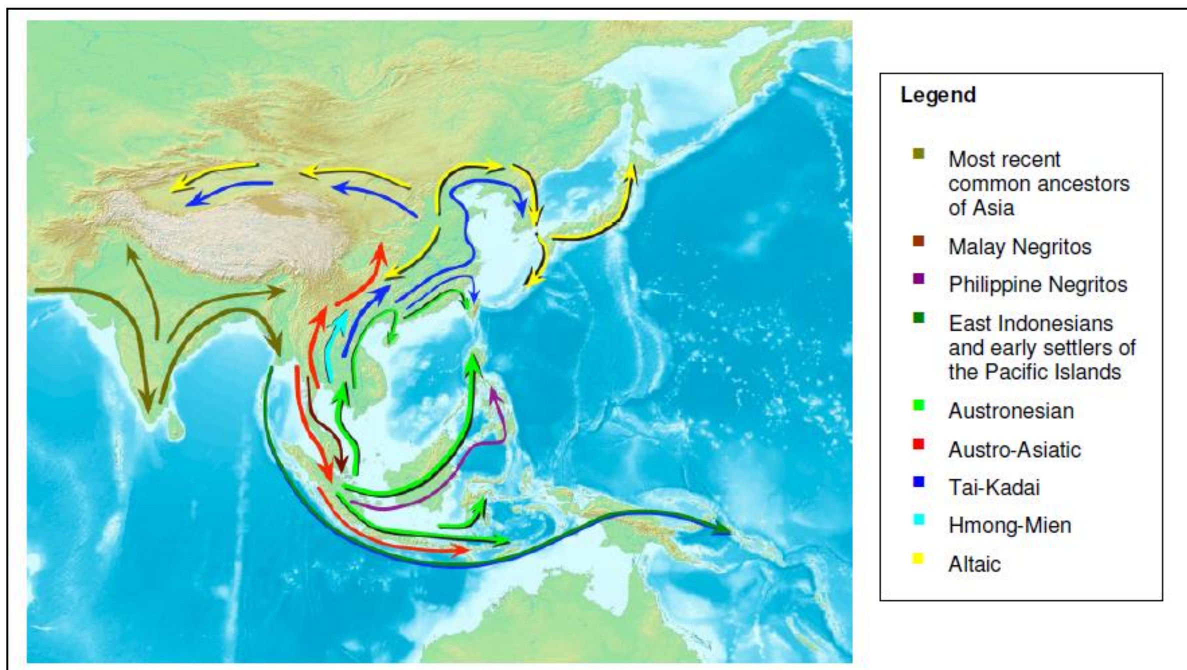


Figure 1-7: Colored arrows show the genetic diversification of humans after they migrated eastward along what is now India's coast and split into numerous genetically distinct groups that moved across Southeast Asia and migrated north into East Asia [Taken from (Normile 2009)].

Above figure shows plausible routes of pre-historical migration of Asian human populations. According to their study, the most recent common ancestors of Asians arrived first in India (aqua-green). Later, some of them migrated to Thailand and South to the lands known today as Malaysia, Indonesia, and the Philippines. The first group of settlers must have gone very far south before they settled successfully. These included the Malay Negritos (brown), Philippine Negritos (purple), the East Indonesians, and early settlers of the Pacific Islands (dark green). Thereafter, one or several groups of people migrated North, mixed with previous settlers there and, finally, formed various populations we now refer to as Austronesian (light green), Austro-Asiatic (red), Tai-Kadai (dark blue), Hmong-Mien (light blue), and Altaic (yellow) etc.

Peninsular Malaysia is a region with a great diversity of human population formed by Malays, Chinese, Indian and the minority aboriginal (Orang Asli) populations. The Orang Asli of Peninsular Malaysia is separated into three main tribal groups, the Negrito, the Senoi and the Proto Malays based on the physical appearance and sociological differences (Hood, 2006). Negrito is believed to be the earliest to arrive in Peninsular Malaysia in abaout 25, 000 years ago. However, to date, Negrito has the least populations among the three Orang Asli groups. The Negritos are physically similar to Andaman Islanders, Aeta in the Philippines, Melanesians and Tasmanians. They were predicted to have originated from Africa and spread throughout Southeast Asia (Macaulay et. al. 2005).

Senoi is the largest Orang Asli group in Peninsular Malaysia and estimated to reach Peninsular Malaysia during the second wave of migration in about 8,000 years ago from South Asia, the mountain areas of Cambodia and Vietnam (Nicholas 1996, Baer 1999). They have Mongoloid physical characteristics and speak Khmer dialects. However, some believe Senoi are descendants of Australoid from Australia and Veddoid from South India (Fix, 1995).

The second largest race of Orang Asli, the Proto-Malays were separated into six tribes, including Jakun/Orang Hulu, Temuan, Semelai, Kuala, Kanaq and Seletar. Fix 1995 had categorized the Proto Malays into 3 categories: tribes who speak Malay and wear Malay

costume (Temuan), tribes with combination of Proto Malays –Senoi from the linguistic and cultural aspects (Senoi) and tribes that settling at coastal areas, mainly Muslim and speaks Sumatran dialects.

Based on archaeology and linguistic findings, it was also suggested that the proto-Austronesian speakers settled in Taiwan about 4000 B.C. before migrated southwards to Southeast Asia region through Philippines into Borneo, Sulawesi, Central Java and Eastern Indonesia 2500 years ago (Andaya 2001).

1.6. OVERVIEW AND AIMS OF THE STUDY

The main aim of this study is to investigate the usefulness of SNP markers as a tool for working with Malaysian casework samples, which typically exhibit high degrees of DNA degradation, due to the high humidity and temperatures.

In order to carry out the study, the potential core panel of 50 to 100 SNP markers needed to be selected. Sanchez and colleagues have reported a multiplex of 52 SNP markers in one PCR reaction with two single base extension reactions (SBE) in the detection of SNPs using CE. The amplicons for PCR ranged from 59 to 115 bp in length. The SBE products ranged from 16 to 92 bp. The assay was shown to be robust using 500 pg DNA input. The two SBE multiplex comprised one reaction with the first 23 primers (marker code 1 to 23) and a second reaction with 29 primers (marker code 24 to 54).

Many studies were carried out that based on these 52 markers (Barbaro et al. 2012, Porras et al. 2009, Santos et al. 2011, Ruiz et al. 2012, Poulsen et al. 2011, Pereira et al. 2008). However, there were also many studies that used 50 markers (Vullo et al. 2011, Schwark et al. 2012a) and 49 markers of these 52 markers (Sharafi Farzad et al. 2013, Borsting et al. 2009, Borsting, Morling 2011, Borsting et al. 2012b, Drobnic et al. 2010, Tomas et al. 2013).

In the 50plex assay, 2 SNPs had to be removed: SNP 7 (rs917118) and SNP 48 (rs8037429) because they led to C artifacts that hampered the evaluation of other SNPs. Schwark and colleagues (2012), generated 2 different multiplex reactions (21- and 29-plex) to obtain a satisfactory detection threshold.

In the 49plex assay, three SNPs (rs1886510/marker 13, rs1463729/marker 35, and rs1028528/marker 53) were excluded. Two of them (marker 13 and marker 53) were previously known to be difficult to analyze and the cause of many extra experiments, because the signals from the SBE products were weak. The SNP rs1463729 (marker 35) was excluded because the peak height of the G allele sometimes was 8 times higher than the peak height of the A allele. This was an unacceptable peak height ratio because the A

allele of a heterozygous individual may not always be detected in weakly amplified samples. In the new design, 49 SNPs are detected in two SBE reactions with 21 and 28 SBE primers, respectively. The lengths of the SBE primers vary from 16 to 84 nucleotides (Borsting et al. 2009).

As for this study, SNP*for*ID 52plex loci were chosen as they represented 52 well characterized SNP loci with population data available for several worldwide populations (Sanchez et al. 2006a). Their study was also supported by the SNPforID Consortium and is one of the sets of SNPs that is recognized by the forensic community (Butler et al. 2008) .

To test these markers the following objectives were defined:

1. Selection of appropriate SNP genotyping/ detection methods.

   i. To assess the sensitivity and reproducibility of two SNP genotyping platforms: TaqMan® SNP Genotyping Assay (Applied Biosystems) using Real-Time PCR and SNaPshot® (Applied Biosystems) single base extension system using ABI PRISM(R) 310 genetic analyzers.

   ii. To optimize and validate both SNP genotyping/detection methods.

2. To develop and optimize a series of multiplexes. In this study, the final four sets of 13-plex SNPs (designated 13_{st}, 13_{nd}, 13_{rd} and 13_{th}) were developed.

3. To genotype and characterize approximately 30 unrelated individuals from each major ethnic group in Malaysia (Malay, Chinese and Indian) using the newly-designed 13-plex assays.

4. To verify and internally validate SNP genotyping with the 13st, 13_{nd}, 13_{rd} and 13_{th} assays on the 3500 Genetic Analyzer (Applied Biosysytems). This was to assess the sensitivity and reproducibility of the assays before application to larger numbers of population samples. This included the use of reduced volumes of reagents and template.

5. SNP genotyping and characterization of an additional 80 samples from each ethnic group.

6. Determine the SNP typing success of problematic specimens. Ten sets of such casework samples were evaluated using the multiplexes designed in order to observe the sensitivity of the assays in generating profiles from degraded DNA samples. In addition, these assays were also applied on 27 bloodstain samples from the same individual but were exposed to the Malaysia's climate (Day 1 to Day 27) sometime around August until September.

7. Perform the multiplex assay for kinship studies. This study used 13 sets of family-related samples obtained from real casework and two control families with known-relationships. This was to assess the multiplex assays' potential as a useful marker set for parentage testing.

8. Statistical analysis and evaluation of the results obtained from the study.

CHAPTER 2

MATERIALS AND METHODS

2.1 Sampling

Throughout the work approval for the use of all Malaysian biological samples was granted by the University Health, Safety and Ethics Committee and the Government of Malaysia (*Appendix A and B*). Informed consents were obtained from the volunteers as mentioned in *Appendix A*. In addition, the research was approved by the University of Central Lancashire's Health Safety and Ethics Committee.

2.1.1. Malaysian population samples

The Malaysian population samples comprised 110 samples from unrelated individuals of each of the three major ethnic groups (Malays, Chinese and Indians) living in Malaysia. Some of these samples have been used in previous studies looking at different forensic markers (Lim et al. 2001, Chang et al. 2002, Seah et al. 2003, Izuan et al. 2005, Chang et al. 2007) These samples, obtained from blood donors in Kuala Lumpur General Hospital, were provided by the Forensic DNA/Serology Laboratory, Department of Chemistry Malaysia. Samples were stored and transported on FTA® cards (Whatman® Bioscience, UK).

2.1.2. Environmental Insult samples

To assess DNA persistence of the blood samples in relation to the environment parameters such as humidity and temperature, 27 portions of cotton material sized 3 cm x 3 cm bearing bloodstains were exposed to the outside environment for 27 days in August 2011 and September 2011. The samples were collected each day (Day 1 to Day 27) and were labelled and then stored immediately at -20 °C until DNA extraction was carried out. The environment parameters/climate for days involved in month of August and September were supplied by the Meteorology Department of Malaysia (*Appendix C*).

2.1.3. Forensic casework samples

A total of 10 cases, which comprised 51 real casework samples (as shown in *Appendix D*) were used to determine the reliability and reproducibility of SNPs assays designed in this study, in generating DNA profiles from the samples. Most of these samples failed to generate full DNA profiles during previous analysis at the DNA Laboratory, Department of Chemistry, Malaysia. The samples were taken from cases that had completed the court trial.

2.1.4. Kinship samples

A total of 12 kinship cases, particularly from parentage testing, were also used to investigate the usefulness and potential of the SNPs assays that had been designed in this study. These samples were also taken from cases that had undergone the court trial. Two volunteer families' samples were used as controls in this study.

2.2. Selection of Single Nucleotide Polymorphisms (SNP) markers

A panel of 52 SNP markers, which were reported to be polymorphic in most populations (Sanchez et al. 2006a) was selected to be used in this study (see Table 2.1 and Table 2.2). These SNP markers were genotyped and validated by The SNP Consortium and a small number by Perlegen and HapMap (http://www.hapmap.org) genotyping initiatives. The chosen SNP markers were selected based on the criteria as mentioned in Sanchez et.al.2006a.

2.3. Quality Control

Essential practices to minimize and monitor contamination within the laboratory were adopted. In all analyses, both positive and negative controls were processed along with the samples.

Table 2-1: 52 SNP markers positions and coordinates (Sanchez et al. 2006a).

Individual Identification SNPs (IISNPs)		**Genome Coordinates**		
Marker Code	**SNP (rs#)**	**Chromosome**	**NCBI human chromosome position**	**Arm locations***
1	rs1490413	1	TSC0724193	p36.32
2	rs876724	2	TSC0208870	p25.3
3	rs1357617	3	TSC0496080	p26.3
4	rs2046361	4	TSC1065282	p16.1
5	rs717302	5	TSC0039610	p15.33
6	rs1029047	6	TSC0253802	p25.3
7	rs917118	7	TSC0229630	p22.2
8	rs763869	8	TSC0065968	p23.3
9	rs1015250	9	TSC0097236	p24.3
10	rs735155	10	TSC0027519	p15.2
11	rs901398	11	TSC0177752	p15.3
12	rs2107612	12	TSC1108144	p13.33
13	rs1886510	13	TSC0904551	q12.11
14	rs1454361	14	TSC0684657	q12
15	rs2016276	15	TSC0326920	q11.2
16	rs729172	16	TSC0028090	p13.3
17	rs740910	17	TSC0105771	p13.2
18	rs1493232	18	TSC0729796	p11.32
19	rs719366	19	TSC0044147	q11
20	rs1031825	20	TSC0334834	p13
21	rs722098	21	TSC0050288	q21.1
22	rs733164	22	TSC0023085	q12.1
23	rs826472	10	TSC0557086	p15.3
24	rs2831700	21	NA	q21.3
25	rs873196	14	TSC0202434	q32.2
26	rs1382387	16	TSC0544547	q23.2
27	rs2111980	12	TSC1113476	q23.2
28	rs2056277	8	TSC1082757	q24.23
29	rs1024116	18	TSC0247167	q23
30	rs727811	6	TSC0062764	q27
32	rs1413212	1	TSC0607362	q43
33	rs938283	17	TSC0357388	q25.3
34	rs1979255	4	TSC0925231	q35.2
35	rs1463729	9	TSC0377760	q33.3
36	rs2076848	11	TSC0022275	q25
37	rs1355366	3	TSC0491662	q28
38	rs907100	2	TSC0186810	q37.3
39	rs354439	13	TSC0700528	q33.2
40	rs2040411	22	TSC1056845	q13.31
41	rs737681	7	TSC0033074	q36.3
42	rs2830795	21	NA	q21.3
43	rs251934	5	TSC0220872	q35.2
44	rs914165	21	TSC0197658	q22.2
45	rs10495407	1	NA	q43.1
46	rs1360288	9	TSC0501229	q33.3
48	rs964681	10	TSC0270699	q26.3
49	rs1005533	20	TSC0082071	q12
50	rs8037429	15	NA	q21.3
51	rs891700	1	TSC0162577	q43.3
52	rs1335873	13	TSC0829150	q12.11
53	rs1028528	22	TSC0253071	q13.31
54	rs1528460	15	TSC0798410	q21.3

Note: *Data were obtained from https://genome-euro.ucsc.edu/cgi-bin.

Table 2-2: PCR and SBE primers used in original study (Sanchez et al. 2006a).

Individual Identification SNPs (IISNPs)		**PCR**			**Single base extension (SBE)**		
Marker Code	**SNP (rs#)**	**Final primer concentration**	**Amplicon size**	**SNP detected**	**Final primer concentration**	**Amplicon size**	**SNPs detected**
1	rs1490413	0.06	68	A/G	0.01	18	T/C
2	rs876724	0.02	83	C/T	0.04	24	C/T
3	rs1357617	0.02	90	A/T	0.02	29	T/A
4	rs2046361	0.03	79	A/T	0.27	78	T/A
5	rs717302	0.04	86	A/G	0.02	74	A/G
6	rs1029047	0.06	100	A/T	0.04	54	A/T
7	rs917118	0.03	87	C/T	0.02	18	G/A
8	rs763869	0.09	100	C/T	0.03	42	C/T
9	rs1015250	0.03	95	C/G	0.02	29	G/C
10	rs735155	0.05	100	A/G	0.02	34	T/C
11	rs901398	0.02	70	C/T	0.04	46	C/T
12	rs2107612	0.05	93	A/G	0.02	58	A/G
13	rs1886510	0.01	86	C/T	0.05	25	G/A
14	rs1454361	0.02	73	A/T	0.02	62	A/T
15	rs2016276	0.17	90	A/G	0.02	74	T/C
16	rs729172	0.02	60	A/C	0.02	70	T/G
17	rs740910	0.08	87	A/G	0.01	42	A/G
18	rs1493232	0.07	59	C/A	0.01	66	T/G
19	rs719366	0.06	105	C/T	0.04	58	C/T
20	rs1031825	0.03	98	A/C	0.02	50	T/G
21	rs722098	0.12	81	A/G	0.01	38	A/G
22	rs733164	0.02	68	A/G	0.01	34	A/G
23	rs826472	0.04	85	C/T	0.01	46	G/A
24	rs2831700	0.04	62	G/A	0.03	56	C/T
25	rs873196	0.03	63	C/T	0.03	52	A/G
26	rs1382387	0.02	69	G/T	0.03	64	G/T
27	rs2111980	0.01	72	A/G	0.03	23	A/G
28	rs2056277	0.03	73	C/T	0.02	92	A/G
29	rs1024116	0.03	76	A/G	0.03	56	A/G
30	rs727811	0.03	78	A/C	0.04	72	A/C
32	rs1413212	0.05	84	A/G	0.05	44	C/T
33	rs938283	0.03	85	C/T	0.05	22	C/T
34	rs1979255	0.03	86	C/G	0.05	40	G/C
35	rs1463729	0.04	87	A/G	0.05	84	A/G
36	rs2076848	0.04	89	A/T	0.05	27	A/T
37	rs1355366	0.05	90	A/G	0.03	48	C/T
38	rs907100	0.04	91	C/G	0.03	27	G/C
39	rs354439	0.05	93	A/T	0.04	80	A/T
40	rs2040411	0.03	94	A/G	0.05	68	A/G
41	rs737681	0.03	96	C/T	0.03	16	C/T
42	rs2830795	0.02	97	A/G	0.03	44	A/G
43	rs251934	0.01	98	C/T	0.03	88	A/G
44	rs914165	0.07	100	A/G	0.05	32	C/T
45	rs10495407	0.05	102	G/A	0.04	68	C/T
46	rs1360288	0.04	103	C/T	0.03	16	A/G
48	rs964681	0.03	106	C/T	0.03	36	C/T
49	rs1005533	0.03	107	A/G	0.03	36	A/G
50	rs8037429	0.03	108	C/T	0.03	76	A/G
51	rs891700	0.04	109	A/G	0.02	32	A/G
52	rs1335873	0.03	110	A/T	0.04	40	T/A
53	rs1028528	0.02	113	A/G	0.03	48	A/G
54	rs1528460	0.06	115	C/T	0.03	60	A/G

Note: SNP markers 31 and 47 were discarded by the authors.

2.4. DNA extraction and purification

The population samples were extracted and purified using three extraction/purification methods; organic solvent and two commercial kits: QIAamp DNA Micro kit and QIAamp DNA Investigator Kit (Qiagen GmbH, Hilden, Germany).

2.4.1. Organic Solvent (Phenol/Chloroform/Isoamyl alcohol)

The organic solvent extraction was carried out as described in previous study (Hochmeister et al. 1991). Approximately, 3 mm^2 of FTA® card (Whatman) was cut and placed into a 1.5 ml microcentrifuge tube. 400 µl of Digest Buffer (10 mM Tris-HCl, 10 mM EDTA, 50 mM NaCl, 2% (w/v) SDS, pH 7.5) and 15 µl of Proteinase K (10 mg/ml; QIAGEN Ltd) were added into the same microcentrifuge tube. The mixture was incubated at 56 °C for 18 h or 24 h. After incubation, 400 µl of phenol:chloroform:isoamyl alcohol (25:24:1; Sigma Aldrich) were added to the mixture, mixed, vortexed and then centrifuged for 5 min at 13,000 rpm (9750 *g)*. The aqueous phase (upper layer) was then transferred into a new 1.5 ml microcentrifuge tube. This step was repeated twice. To each tube, 1 ml of cold absolute ethanol was added. The tubes were inverted few times and then incubated at -20 °C for 1 h. After that, the tubes were centrifuged for 20 min at 13,000 rpm. Ethanol was removed by pipetting out from the tubes. The tubes caps were left open to allow any residual ethanol to evaporate. DNA was re-suspended by adding 30 µl of TE buffer (10mM Tris-HCL, 0.1mM EDTA, pH 8.0) and was incubated at 56 °C overnight.

2.4.2. QIAamp DNA Micro kit

The QIAamp DNA Micro kit was used by following the manufacturer’s protocol (QIAGEN 2010b). This protocol is for isolation of genomic DNA from blood card punches (diameters of approximately 3 mm), bloodstain card or FTA Card (Whatman), Guthrie test cards and untreated blood or blood treated with anticoagulants such as EDTA, citrate or heparin. Approximately, up to three 3 mm diameter punches from a

dried blood spot were placed into a 1.5 ml microcentrifuge tube. To the tube, 180 µl of Buffer ATL and 20 µl of Proteinase K (10 mg/ml) were added into the tube. The mixed solution was vortexed thoroughly. The tube was placed in the heated incubator at 56 °C for 1 h. During incubation, the solution was vortexed vigorously for 10 s every 10 min. After the incubation, the tube was briefly centrifuged to remove drops from the inside of the lid. Then, 200 µl of Buffer AL was added and the lid was closed before mixed by pulse-vortexing for 10 s. The tube was incubated again at 70 °C for 10 min and vortexed for 10 s every 3 min. The tube was centrifuged briefly before transferring the entire lysate into the QIAamp MinElute column without wetting the rim. The tube was centrifuged at 8,000 rpm (6000 *g)* for 1 min. Then, the QIAamp MinElute column was placed in a clean 2 ml collection tube, whilst the collection tube containing the flow-through was discarded. 500 µl of Buffer AW1 was added to the column, and the tube was centrifuged at 8,000 rpm for 1 min. The filtrate was discarded, and the column was placed in a clean collection tube, before 500 µl of Buffer AW2 was added into the column. The tube was centrifuged again at 8,000 rpm for 1 min.

The column was transferred to a clean collection tube and the flow-through solution was discarded. Then the column was centrifuged at full speed (13,000 rpm or 9750 *g*) for 3 min to completely dry the column's membrane. The QIAamp MinElute column was transferred in a clean 1.5 ml microcentrifuge tube and the filtrate was discarded. To each tube, 25 µl of Buffer AE was added to the centre of the column's membrane, and the column was incubated at room temperature for 1 min. After that the tube was centrifuged at higher speed (13,000 rpm) for 1 min. The filtrate containing the DNA was stored at -20 °C.

2.4.3. QIAamp DNA Investigator Kit

The QIAamp DNA Investigator kit was used by following the manufacturer's protocol (QIAGEN 2010a). Firstly, stained materials were cut out up to 0.5 cm^2, then into smaller pieces and transferred to a 1.5 ml microcentrifuge tube. To the tube, 300 µl of Buffer ATL and 20 µl of Proteinase K were added, and the solution was mixed by pulse-vortexing for 10 s. The tube was kept in the incubator at 56 °C for 1 h. The solution was

vortexed for 10 s every 10 min to enhance the lysis process. After the incubation, the tube was centrifuged briefly, and 300 µl of Buffer AL was added. The solution was mixed by pulse-vortexing for 10 s, and placed it in the incubator at 70 °C for 10 min. Every 3 min, the solution was vortexed for 10 s. Then, the tube was centrifuged briefly, before adding 150 µl of ethanol (96%), and mixed thoroughly by pulse-vortexing for 15 s. The supernatant was transferred to the QIAamp MinElute column, and was centrifuged at 8,000 rpm (6,000 *g)* for 1 min. The column was placed in a clean collection tube, whilst the flow-through was discarded. To the column, 500 µl of Buffer AW1 was added and the tube was centrifuged at 8000 rpm for 1 min. The filtrate was discarded, the column was placed in a clean collection tube, and 700 µl of Buffer AW2 was added into the column. The tube was centrifuged again at 8000 rpm for 1 min. The column was transferred in a clean collection tube, and the flow-through was discarded. After that, 700 µl of ethanol (96%) was added and the tube was centrifuged at 8000 rpm for 1 min. The column was transferred again in a clean 2 ml collection tube and was centrifuged at full speed (13,000 rpm) for 3 min. The QIAamp MinElute column was placed in a clean 1.5 ml microcentrifuge tube, whilst the flow-through was discarded. Finally, 25 µl of Buffer ATE was dispensed in the centre of the membrane and left to incubate at room temperature for 1 min. Then, the tube was centrifuged at full speed (13, 000 rpm) for 1 min. DNA was stored until being used.

2.5. Quantification of DNA samples

The DNA concentrations were determined using Quantifler® Human DNA Quantification kit (Applied Biosystems) run on an Applied Biosystems® 7500 Real-Time PCR System. Quantification standards were prepared by serial dilution in TE buffer of the 200 ng/µl Human DNA Standard included in the kit. Eight serial dilutions ranging from 50 ng/µl to 0.023 ng/µl were used. The procedure was carried out according to the manufacturer’s protocol, but instead of using full-scale reactions, the total volume of this assay was reduced (Westring et al. 2007). The reduced-scale reaction consisted of a 5.25 µl of Quantifiler PCR Reaction Mix, 6.25 µl of Quantifiler Human Primer Mix and 1 µl of the DNA sample for a final reaction volume of 12.5 µl.

Samples, including the non-template control (NTC) and the DNA standards were loaded into MicroAmp® Optical 96-well reaction plate and sealed using Optical Adhesive Covers (Applied Biosystems). Each standard dilution was run in duplicate, with replicates placed in wells near each other, then followed by DNA samples.

The sample sheets were set up and the samples were run using the standard thermal amplification profile of 95 °C for 10 min; 40 cycles of 95 °C for 15 s and 60 °C for 1 min. Standard curves were generated and amplification plots were used to compare the results. Internal Positive Control (IPC) results were monitored for the presence of PCR inhibitors.

From the results obtained, the samples with low DNA concentrations (< 0.01 ng) were re-quantified to confirm no error had happened during the sample preparation. The samples which still showed very low DNA amounts were re-extracted from the original FTA card using the Organic Solvent (phenol/chloroform/isoamyl alcohol) method and then quantified with Quantifler® kit.

2.6. Reproducibility and Sensitivity study

In order to evaluate the ability of the SNP genotyping method chosen for obtaining reliable and consistent results, a small internal validation study was performed based on the threshold level of 2 ng genomic DNA, which is now above the sensitivity level used in forensic casework.

2.6.1. Serial dilutions

A series of dilutions of Applied Biosystems TaqMan® Control Genomic DNA (Human; 10 ng/ µl) were prepared as follows: (i) TaqMan® SNP Genotyping assays : (2.0, 1.0, 0.5, 0.25, 0.125, 0.063, 0.032, 0.016, 0.008, 0.004 and 0.002) ng/ µl, and (ii) SNaPshot Reaction *(multiplex)*: (2.5, 1.0, 0.5, 0.25, 0.13, 0.06, 0.03, 0.02 and 0.01) ng/ µl.

2.6.2. Testing population samples

1 µl of each extracted DNA sample (Malaysian population samples), which ranged from 0.01 ng/ µl to 569 ng/ µl was used to assessed the reproducibility of the four sets of SNaPshot multiplexes designed in this study.

2.7. SNP Genotyping

In this study, two SNP typing technologies were selected to genotype the autosomal SNP markers; (i) TaqMan® SNP Genotyping Assays and (ii) SNaPshot® Multiplex System (both, Applied Biosystems). The SNP typing was carried out using the AB® 7500 Real-Time PCR System and Genetic Analyzer 310 (both Applied Biosystems), respectively.

2.7.1. TaqMan® SNP Genotyping Assays

2.7.1.1. Concept of the assay

The assay of these SNP markers contains two locus-specific primers that flank the SNP of interest and two TaqMan® MGB probes. One probe labelled with VIC® dye detects the Allele 1 sequence and another probe labelled with FAM™ dye detects the Allele 2 sequence. The assay determines the presence or absence of the two possible variant alleles at the SNP site in a DNA target sequence based on the change in fluorescence of the dyes associated with the probes.

2.7.1.2. Sample preparation

Firstly, the TaqMan Genotyping Assay mix (consisting of primers and probe) were diluted from 40X to a 20X working stock with 1X TE buffer, vortexed and centrifuged. The assay was aliquotted and stored at -20 °C until it was used. Then, a 96-well

reaction plate for an Allelic Discrimination (AD) assay was set up with the samples and reaction mix. The optimization was carried out using TaqMan® Control Genomic DNA (Human; 10 ng/ µl) with different series of concentrations, and different PCR reaction mix volumes as follows:

Table 2-3: Several PCR reaction mix volumes used for the study to observe the sensitivity of the assay.

PCR Reaction Mix Volume (µl/well)				
Components	***5µl reaction***	***10µl reaction***	***12.5µl reaction***	***25µl reaction***
TaqMan Genotyping Master Mix (2X)	2.5	5	6.25	12.5
TaqMan genotyping Assay Mix (2X)	0.25	0.5	0.625	1.25
DNase-free, RNase free Water	1.25	3.5	4.625	10.25
Total	**4.0**	**9.0**	**11.5**	**24.0**

The reaction mix was pipetted into each well of a reaction plate according to the calculated volumes as shown in Table 2.3. Then, 1 µl from each series of dilution was added into each well. The plate was covered with MicroAmp® Optical Adhesive Film and each well was inspected for volume uniformity or bubbles before putting it into the AB® 7500 Real- Time PCR instrument.

2.7.1.3. Creating an Allelic discrimination (AD) Plate Document

An AD plate document was created to store all information about the run, including sample names, markers, and detectors following the protocols provided by the manufacturers (Applied Biosystems 2005).

2.7.1.4. Creating an Absolute Quantification (AQ) Plate Document

An AQ plate document was created to store real-time data for AD assays following the protocols provided by the manufacturers (Applied Biosystems 2005). This document was used only to amplify target sequences (not to quantify the PCR products), and therefore no standard curves were needed.

2.7.1.5. Performing Runs

The AD assay involved three main steps (provided by the manufacturer in the User Manual (Applied Biosystems 2005)) as follows in order to generate data and results:

2.7.1.5.1. A pre-read run

A run on an AD plate document to determine the baseline fluorescence associated with primers and probes before amplification.

2.7.1.5.2. An amplification run

A run uses an Absolute Quantification (AQ) plate document to generate real-time PCR data, which can be used to analyze and troubleshoot the PCR data for the AD assay. The cycling program used; 50 °C for 2 min; 1 cycle of 95 °C for 10 min; and 40 cycles of 95 °C for 15 s and 60°C for 1 min.

2.7.1.5.3. A post-read run

A run that uses the original AD plate document (pre-read plate document). The post-read run automatically subtracts the baseline fluorescence determined during the pre-read run, and then assigns allele calls (automatically or manually) using the amplified data.

All above steps were carried out according to the manufacturer's protocol.

2.7.1.6. Evaluating results

After an AD post-read run, the 7500 SDS Software was used to analyze raw data. The raw data were converted and expressed in terms of fluorescence signal versus filters, to pure dye component. The software identified the dye components and determined the contribution of each dye in the raw data using the multicomponent algorithm. The results of AD were plotted on a scatter plot of Allele X versus Allele Y.

2.7.2. SNaPshot® Multiplex System

52 pairs of SNPs primers (Sanchez et al. 2006a) were used in single and multiplex SNaPshot reactions following a flow diagram as shown in Figure 2-1.

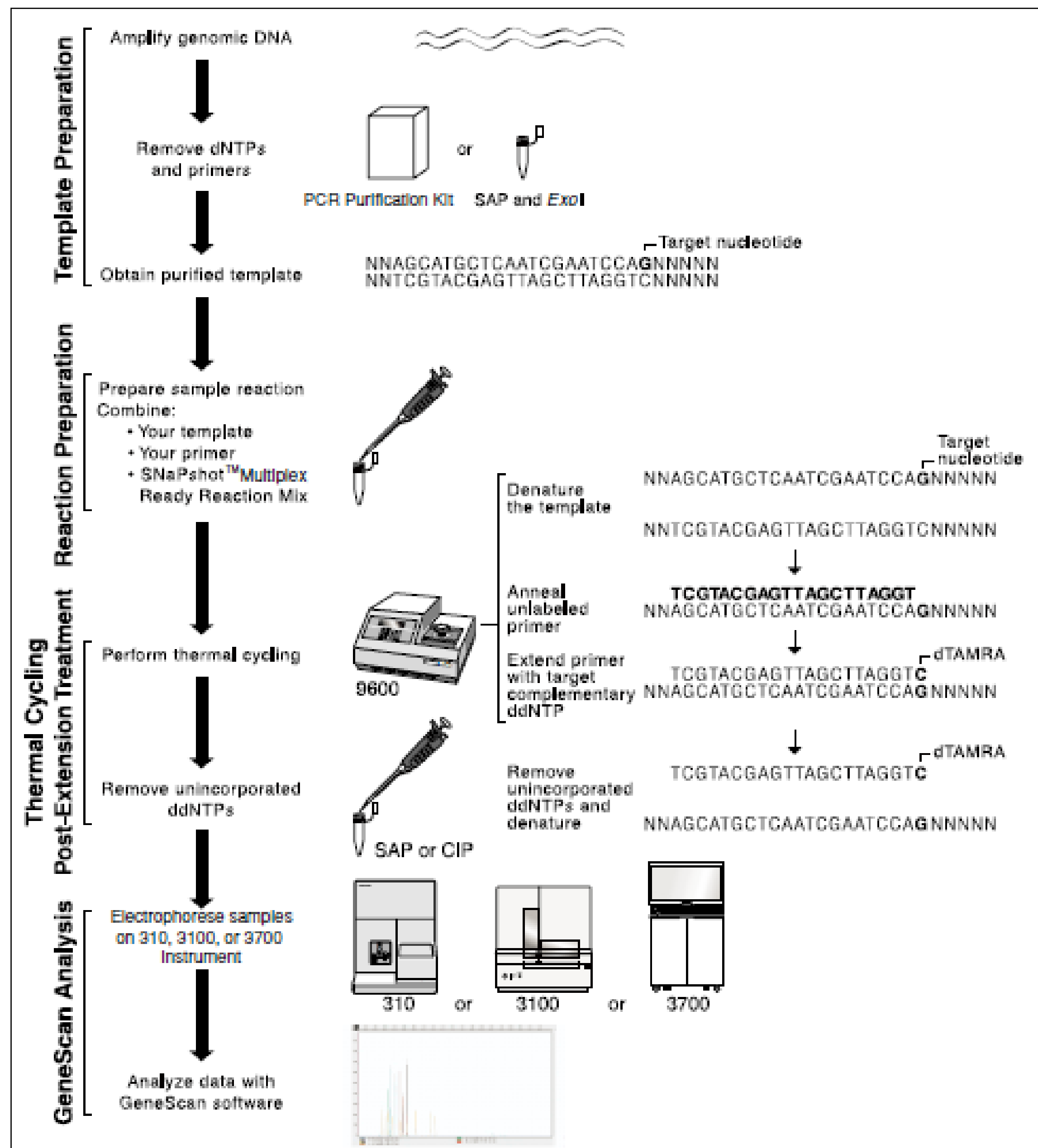


Figure 2-1: Overview of the procedure of PCR and SBE reactions for SNP genotyping whole process [Taken from Applied Biosystems 2009a].

2.7.2.1. Single SNP reaction

Each of the 52 SNP markers was amplified individually in a single PCR reaction. Two control DNA samples (TaqMan® Control Genomic DNA and one Malay individual (M7-7)) were used throughout the optimization study. Each was used at two different DNA concentrations: 0.5 ng and 1 ng to test the assays. Several concentrations of PCR and single base extension (SBE) primers were evaluated to determine the sensitivity of the assay (Table 2.4).

Table 2-4: Combinations of primer concentrations in SNaPshot analysis.

Stage	Combination of the primer concentrations (µM)					
PCR	0.05	0.05	0.1	0.1	0.2	0.2
SBE	0.05	0.1	0.05	0.1	0.05	0.1

2.7.2.2. Multiplex SNP assay set-up

In this study, four sets of 13-plexes were developed, named as 13_{st}, 13_{nd}, 13_{rd} *and* 13_{th}. These multiplexed assays were used and tested for the optimal conditions. All sequences and sizes of all primers used are the same as described in the original study (Sanchez et al. 2006a). The combinations of the PCR and SBE primers were evaluated to increase the specificity of the assays. The final combinations of all primer pairs in each multiplex (PCR and SBE reactions) are presented in Table 2.5.

2.7.2.3. Amplification and optimization of SNP assays

Optimization of each reaction was carried out to determine the optimal amplification conditions for different DNA targets. Two sets of reaction volumes, 12 µl and 15 µl were prepared as listed in Table 2.6. The final concentrations of each primer (forward and reverse) in single reaction was 0.2 µM, whilst for 13-plex reactions, it ranged

from 0.05 µM to 0.4 µM as shown in Table 2.7. All PCR were performed in a GeneAmp 9700 or Veriti™ 96 Well thermal cycler (both, Applied Biosystems), with the optimal cycling parameters (Table 2.8).

2.7.2.4. Agarose Gel Electrophoresis

The amplified products (5 µl) of single SNP reactions were analyzed using a 2.5% (w/v) agarose gel which was stained afterwards with ethidium bromide (0.5 µg/ml). The 20 bp and 1 kb DNA Marker were used and loaded together with the samples onto the gel. The gel was photographed using a gel documentation system (Bio Doc-It™ system at 595 nm).

Table 2-5: New combination and arrangement of the same 52 SNP code markers (from the original study) into four sets of 13-plex assays, *13_{st}, 13_{nd}, 13_{rd} and 13_{th}.*

Four sets of 13-plex PCR/SBE assays			
Markers Code			
13st	**13nd**	**13rd**	**13th**
1	*46*	*41*	*7*
2	*13*	*27*	*33*
3	*9*	*36*	*38*
22	*10*	*21*	*51*
34	*49*	*44*	*48*
23	*11*	*8*	*52*
53	*4*	*17*	*42*
20	*25*	*32*	*37*
12	*29*	*6*	*54*
14	*24*	*19*	*45*
16	*26*	*40*	*50*
5	*18*	*30*	*35*
39	*15*	*43*	*28*

Table 2-6: PCR optimal conditions of single and multiplex reactions.

Final concentrations (volume used in µl)		
Components	***Single reaction***	***Multiplex***
Concentration of DNA used	*0.5 ng/ µl and 1 ng/ µl(1 µl)*	*0.01 ng/ µl to 32 ng/ µl (1 µl)*
Primers (Forward and reverse)	*0.2 µM each (0.96 µl)*	*0.05 µM to 0.3 µM (1 µl)*
10X PCR Buffer II (100 mM Tris-HCL, 500 mM KCl, pH 8.3)	*1X (1.2 µl)*	*1X (1.5 µl)*
25 mM $MgCl_2$ Solution	*1.5 mM (0.72 µl)*	*3.0 mM (1.8 µl)*
10 mM GeneAmp® dNTPs	*0.2 mM (0.24 µl)*	*0.4 mM (0.6 µl)*
AmpliTaqGold Polymerase, 5U/µl	*1.25U (0.25 µl)*	*2U (0.4 µl)*
DNase-free, RNasefree Water	**7.63 µl*	**8.7 µl*
Total Volume	***12 µl***	***15 µl***

*Any volume of water can be used as long as the total volume of the Master Mix, sample and primers equal to 12 µl or 15 µl.

Table 2-7: Optimized final primer concentrations (µM) for each multiplex set.

Set of Multiplexes	13st		13nd		13rd		13th	
Multiplexed Loci with final primers concentrations (µM)	1	*0.15*	46	*0.1*	41	*0.1*	7	*0.1*
	2	*0.1*	13	*0.1*	27	*0.05*	33	*0.08*
	3	*0.15*	9	*0.1*	36	*0.1*	38	*0.15*
	22	*0.1*	10	*0.15*	21	*0.15*	51	*0.1*
	34	*0.15*	49	*0.1*	44	*0.1*	48	*0.15*
	23	*0.2*	11	*0.15*	8	*0.15*	52	*0.15*
	53	*0.1*	4	*0.3*	17	*0.1*	42	*0.1*
	20	*0.15*	25	*0.1*	32	*0.3*	37	*0.15*
	12	*0.15*	29	*0.1*	6	*0.25*	54	*0.15*
	14	*0.1*	24	*0.1*	19	*0.2*	45	*0.4*
	16	*0.1*	26	*0.15*	40	*0.1*	50	*0.2*
	5	*0.15*	18	*0.1*	30	*0.15*	35	*0.1*
	39	*0.1*	15	*0.15*	43	*0.15*	28	*0.07*

Table 2-8: The optimal thermal cycling parameters.

PCR Cycling conditions	Single reaction	Multiplex reaction
Initial denaturation	*95 °C for 1 min*	*95 °C for 5 min*
Denaturation Annealing Extension	*95 °C for 30 sec* } *35X* *60 °C for 30 sec* } *65 °C for 30 sec*	*95 °C for 30 sec* } *35X* *60 °C for 30 sec* } *65 °C for 30 sec*
Final extension	*65 °C for 7 min*	*65°C for 7 min*

2.7.2.5. PCR product Clean-up

When amplification was complete, any non-incorporated dNTPs and primers remaining were removed using the ExoSAP-IT®, a single step enzymatic reaction (USB, Affymetrix). In each 3 µl PCR samples, 1.2 µl of ExoSAP-IT® reagent was added and then incubated at 37 °C for 15 min and 80 °C for another 15 min. The samples were kept in -20 °C until the single base reaction (SBE) amplification was carried out.

2.7.2.6. Single Base-reaction (SBE) Amplification

To develop the four sets of 13 SBE multiplexes in this study a few of the SBE primers length were modified from the original SBE primers established (Sanchez et al. 2006a). Each modification involved the addition or reduction of poly thymidine (poly T) or poly cytosine (poly C) to the 5' end of SBE primers, as suggested by the Applied Biosystems SNaPshot® User's Manual. The new SBE length (bp) is shown in Table 2.9.

The SBE reactions were carried out for both, single reaction and multiplex reaction using SNaPshot® Multiplex Kit (Applied Biosystems). The optimized reactions and optimized SBE primers (single and multiplex) are illustrated in Table 2.10 and Table 2.11, respectively. Thermal cycling conditions were applied as suggested in the SNaPshot™ protocol: 25 cycles of 96 °C for 10 s, 50 °C for 5 s, and 60 °C for 30 s.

Table 2-9: Modified SBE length (bp).

SNP Marker Code	Original length (bp)	Modified length (bp)	SNP detected
4	*78*	*47*	*T/A*
5	*74*	*70*	*A/G*
7	*18*	*20*	*G/A*
8	*42*	*40*	*C/T*
9	*29*	*30*	*G/C*
11	*46*	*44*	*C/T*
13	*25*	*26*	*G/A*
16	*70*	*68*	*T/C*
18	*66*	*68*	*T/G*
19	*58*	*60*	*C/T*
21	*38*	*32*	*A/G*
22	*34*	*36*	*A/G*
23	*46*	*45*	*G/A*
51	*32*	*31*	*A/G*

Table 2-10: The optimal conditions used for the SBE reactions.

Components	Volumes (µl)			
	Single reaction	Multiplex reaction	Positive Control	NTC Control
SNaPshot® Mix	2	2	2	2
Purified PCR product/ Control DNA	1	1.5	1	1
SBE primer (0.2 µM) / SBE primers/ Control primers	0.2	1	0.5	0.5
DNase-free, RNasefree Water	1.8	0.5	1.5	1.5
Total	5	5	5	5

Table 2-11: The final concentrations (µM) of SBE primers used in the SNaPshot reactions.

SNaPshot Reactions	Single reaction	Multiplex reaction							
		13st		13nd		13rd		13th	
Multiplexed Loci with final primers concentrations (µM)	Each single SNP reaction used 0.2 *µM* *SBE* primer	1	*0.07*	46	*0.07*	41	*0.07*	7	*0.15*
		2	*0.1*	13	*0.06*	27	*0.05*	33	*0.1*
		3	*0.08*	9	*0.15*	36	*0.08*	38	*0.15*
		22	*0.1*	10	*0.15*	21	*0.15*	51	*0.1*
		34	*0.15*	49	*0.1*	44	*0.15*	48	*0.15*
		23	*0.2*	11	*0.1*	8	*0.2*	52	*0.1*
		53	*0.1*	4	*0.15*	17	*0.15*	42	*0.15*
		20	*0.2*	25	*0.15*	32	*0.3*	37	*0.15*
		12	*0.3*	29	*0.1*	6	*0.4*	54	*0.15*
		14	*0.15*	24	*0.15*	19	*0.2*	45	*0.2*
		16	*0.2*	26	*0.1*	40	*0.1*	50	*0.15*
		5	*0.3*	18	*0.2*	30	*0.15*	35	*0.15*
		39	*0.2*	15	*0.3*	43	*0.2*	28	*0.1*

2.7.2.7. SBE products Clean-up by SAP (USB, Affymetrix)

The excess of fluorescently labeled ddNTPs in the primer extension reaction were removed by the addition of 0.5 µl Shrimp Alkaline Phosphatase (SAP) into each 2.5 µl SBE product, the reactions were mixed briefly and incubated at 37 °C for 45 min and 80 °C for 15 min.

2.7.2.8. Capillary Electrophoresis

2.7.2.8.1. ABI PRISM® 310 Genetic Analyzer (Applied Biosystems)

Electrophoresis running buffer was used at 1X concentration. The GSPOP 4 (1ml) E.md5 run module with dye set DS-02 (filter set E): dR110, dR6G, dTAMRA™, dROX™, and LIZ® dyes) was used with the following parameters: run temperature 60 °C, syringe pump time 150 s, pre-run voltage 15 kV, pre-run time 120 s, injection time 5 s, injection voltage 15 kV, run voltage 15 kV and run time 30 min.

In a 0.2 ml PCR tube, 1 µl of SAP- treated SBE products was diluted in 11 µl of Hi-Di™ formamide and 0.4 µl GeneScan™ (GS) 120-LIZ® Internal Size Standards (both, Applied Biosysytems). Then, the samples were analyzed using ABI 310 PRISM® Genetic Analyzer (Applied Biosystems) with 36 cm capillary array and POP-4™ polymer.

2.7.2.8.2. 3500 Genetic Analyzer (Applied Biosystems)

For the concordance study and to increase the throughput of sample processing, DNA fragment analysis was also performed using a 3500 Genetic Analyzer. Following PCR amplification, electrophoretic separation of dye-labelled amplicons was performed on the 3500 using a 50 cm capillary array and 3500 POP-6™ polymer. The FragmentAnalysis50_POP6 run module was used in combination with the dye set DS-02 (filter set E5): dR110, dR6G, dTAMRA™, dROX™, and LIZ® dyes) with the following parameters: run temperature 60 °C, pre-run voltage 15 KV, pre-run time 180 s, injection time 10 s, injection voltage 1.6 kV, run voltage 15 kV and run time 1800 s. Each sample for fragment analysis was prepared by adding 1 µl of PCR product to 10 µl of Hi-Di™ formamide (Applied Biosystems) containing 0.3 µl GS120 LIZ size standard (Applied Biosystems). The samples were heated at 95 °C for 5 min, and snap-cooled at least 3 min. Samples were subjected to capillary electrophoresis, detected by laser-induced fluorescence.

2.7.2.8.3. Analysis of DNA Profiles

The data obtained from capillary electrophoresis (CE) were analysed using GeneMapper™ ID version 3.2 (ABI 310) and GeneMapper® *ID-X* Software version 1.2 (3500 Genetic Analyzer). The parameters for the analysis of the multiplex amplicons were kept constant for each run and are described in Table 2.12.

Table 2-12: Shown below are the parameters for the analysis of PCR fragments.

Parameters	Values
Analysis Range	Full Range
Size Call Range	All Sizes
Size Calling Method	Local Southern
Peak Detection	50 RFU
Baseline Window	51 pts
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. Data collection and evaluation.

All SNaPshot results were collected and analyzed with peak thresholds set to a minimum of 120 RFUs (blue colour), 60 RFUs (green colour) and 30 RFUs (yellow, red and orange colours). The criteria for acceptable peak height ratio for heterozygous allele calls and homozygous allele calls for each locus were as described in the original study (Sanchez et al. 2006a).

2.9. Statistical analysis

The results were further analyzed using Arlequin Software version 3.5.1.3 (Excoffier, Lischer 2010) for allele frequencies, expected and observed heterozygosities (*He* and *Ho*, respectively), Hardy-Weinberg equilibrium (HWE) and pairwise F_{ST} values. The analyses Arlequin can perform on the data fall into two main categories: intra-population and inter-population methods. In the first category statistical information is extracted independently from each population, whereas in the second category, samples are compared to each other.

HWE is a test of non-random association of alleles within diploid individuals. It concerns the relationship between allele probabilities and genotype probabilities at one locus. In essence, the Hardy–Weinberg law is a statement of independence between alleles at one locus.

Estimates of F_{ST} can identify regions of the genome that have been the target of election, and comparisons of F_{ST} from different parts of the genome can provide insight into the demographic history of those populations. In Wright's notation, subscripts *ST refer* to "subpopulations within total". F_{ST} plays a central role in population and evolutionary genetics, and F_{ST} has wide applications in fields from disease association mapping to forensic science. Thus, F_{ST} must be considered when allele frequencies are compared between "cases" and "controls" to ensure that differences between them are greater than expected by chance. Similarly, the match probability between a suspect and a crime scene sample is specific to the set of people who might reasonably be expected to be sources of the sample. But defining this set is difficult, so a "θ

correction" is applied to population frequencies to accommodate variation among subpopulations. The "θ correction" depends on the value of *FST*. θ, the co-ancestry coefficient, is the probability that two alleles in the population have descended from the same allele and are identical by descent. This is a measure of the coancestry of populations diverging due to genetic drift. The larger θ, the longer it has been since the populations diverged.

The power of discrimination and match probability values were calculated for each locus using Powerstats software (Jones 1972, Botstein et al. 1980).

CHAPTER 3

DEVELOPMENT, OPTIMIZATION AND VALIDATION OF THE SNP DETECTION ASSAYS

PART ONE:
TAQMAN® SNP GENOTYPING ASSAYS

PART TWO:
ANALYSIS OF SNaPShot ASSAYS USING ABI PRISM® 310 GENETIC ANALYZER

PART ONE:

TAQMAN® SNP GENOTYPING ASSAYS

3.1. INTRODUCTION

The analysis of degraded DNA can be problematic. As DNA becomes degraded, the higher molecular weight STR loci fail to amplify (Golenberg et al. 1996) giving a partial DNA profile that has a lower discrimination power (Dixon et al. 2005). Recent advances in the identification and analysis of single nucleotide polymorphisms (SNPs) have demonstrated the advantage of these markers over short tandem repeats (STRs) is that they only require small amplicons (Goodwin et. al. 2011).

There are various platforms for detecting SNP markers in single or multiplex reactions. TaqMan® SNP Genotyping Assays are the largest collection of single-tube and ready-to-use SNP assays that are commercially available. The TaqMan SNP® Genotyping Assays library (snp.appliedbiosystems.com) consists of four human assay collections:the TaqMan® Validated SNP Genotyping Assays, TaqMan® Coding SNP Genotyping Assays, TaqMan® Pre-Designed SNP Genotyping Assays, and TaqMan® Drug Metabolism Genotyping Assays (Applied Biosystems 2006).

The TaqMan® SNP Genotyping assays are a single-tube assay that exploit the 5′ exonuclease activity of AmpliTaq Gold® DNA Polymerase to discriminate the targeted alleles. The assay comprises two locus-specific PCR primers that flank the SNP of interest, and two allele-specific oligonucleotide TaqMan® probes (Figure 3-1). These probes have a fluorescent reporter dye at the 5′ end, and a non-fluorescent quencher (NFQ) with a minor groove binder (MGB) at the 3′ end (Applied Biosystems 2010a).

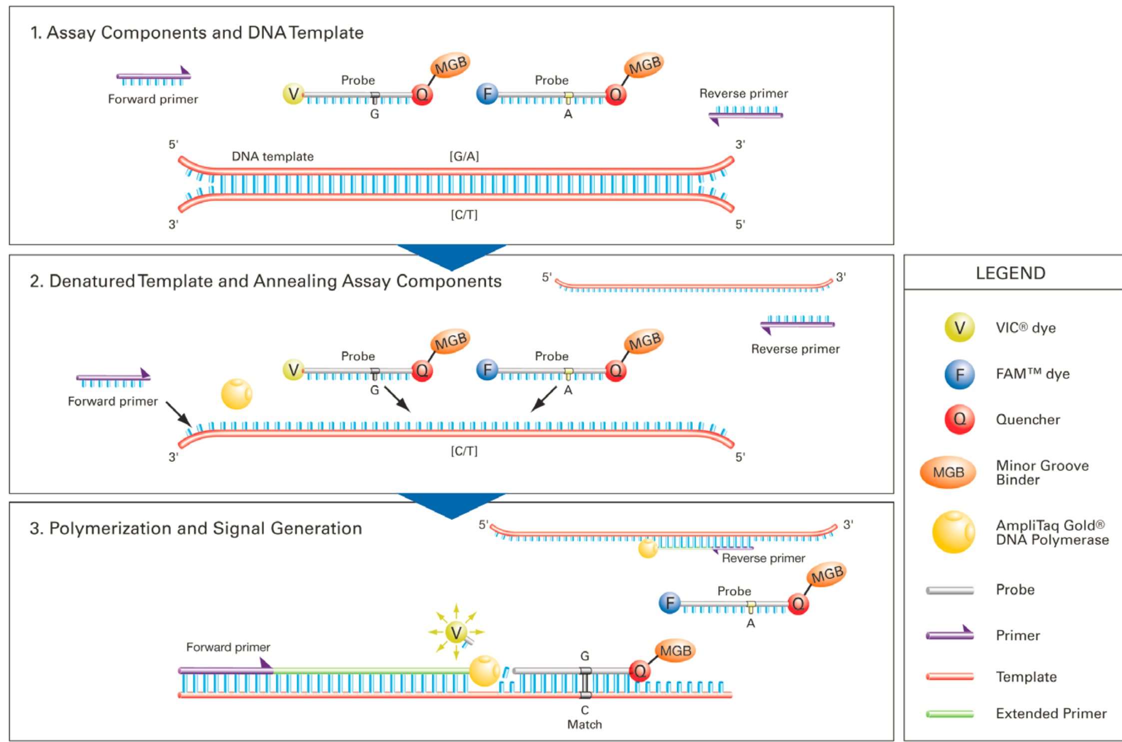


Figure 3-1: Allelic discrimination is achieved by the selective annealing of TaqMan® MGB probes. [Taken from Applied Biosystems, 2010a].

For each sample in an allele discrimination (AD) assay, a unique pair of fluorescent dye detectors is used. The AD assay measures the change in fluorescence of the dyes associated with the probes. One fluorescent dye is a perfect match to the wild type (allele 1) and the other fluorescent dye detector is a perfect match to the variant (allele 2) (Applied Biosystems 2010a). A large number of the validated TaqMan® SNP Genotyping assays are commercially available, which makes it possible to implement this type of SNP genotyping platform (De La Vega et al. 2005b).

As for data analysis, TaqMan® Genotyper Software is commonly used as a SNP genotyping data analysis tool for TaqMan® SNP Genotyping Assays. This software is a standalone application that analyzes raw data from genotyping experiments performed on an Applied Biosystems® realtime PCR system. The software allows the user to choose between two methods to call genotypes: Auto-calling and Classification Scheme. In the Auto-calling method, the determining factor influencing the genotype call is a value that the algorithm assigns to each data point; that is the Quality value (Life Technologies 2013).

The Quality value reflects the probability of the genotype call; the default threshold for the Quality value in the software is 0.95. This threshold is listed in the Analysis Settings on the QC Settings tab (Life Technologies 2013). Lowering the default threshold can increase the sample call rate (Figure 3-2 (B)), but usually a sample with a lower Quality value of 0.94, will receive an undetermined call (Figure 3-2 (A)).

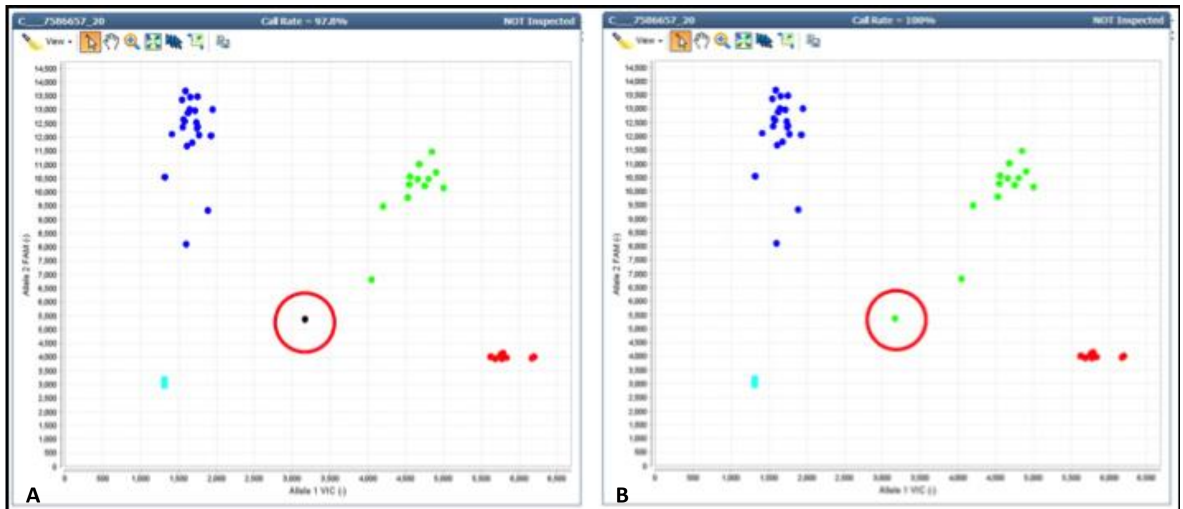


Figure 3-2: The circled sample in the scatter plot (A) is assigned an undetermined call when the default threshold of 0.95 is applied for analysis, but it is assigned a heterozygous call (B) when the Quality value threshold is lowered to 0.80. [Taken from Life Technologies (2013)].

Referring to Figure 3-3, Angle (A) and amplitude (B) for each data point are calculated by the formulas below (Life Technologies 2013).

$$\text{Angle} = \tan^{-1} \frac{\text{FAM Rn} - \text{mean NTC FAM Rn}}{\text{VIC Rn} - \text{mean NTC VIC Rn}}$$

$$\text{Amplitude} = \sqrt{(\text{FAM Rn} - \text{mean NTC FAM Rn})^2 + (\text{VIC Rn} - \text{mean NTC VIC Rn})^2}$$

The algorithm computes the models for the distribution of angles and amplitudes for each of the class labels. The score is an estimate of how closely a given data point belongs to those underlying models of angle and amplitude distributions for each genotype cluster.

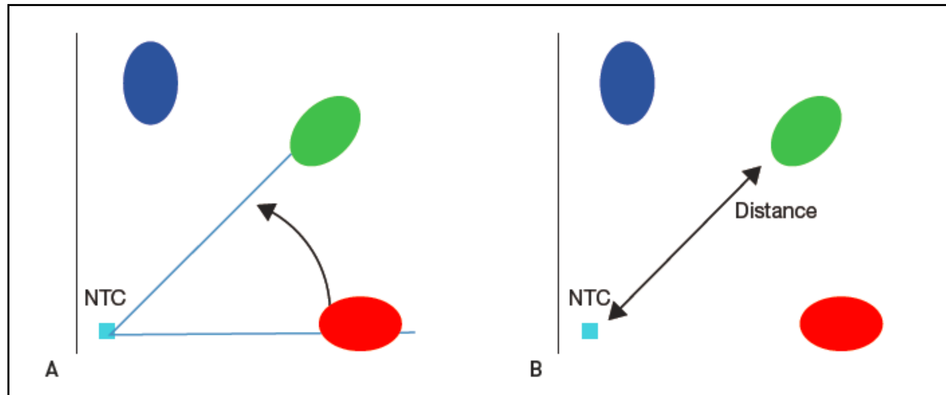


Figure 3-3: Illustration of angle (A) and amplitude (B) calculated from the No Template control (NTC). Good cluster separation means good separation of data points from other clusters (angle) and leads to good angle scores. Good signal intensity means good separation of data points from the NTC and generates good amplitude scores. [Taken from Life Technologies (2013)].

The distribution of angles and amplitudes shown in Figure 3-4, shows that the data point would map between the distributions of NTC and genotype clusters.

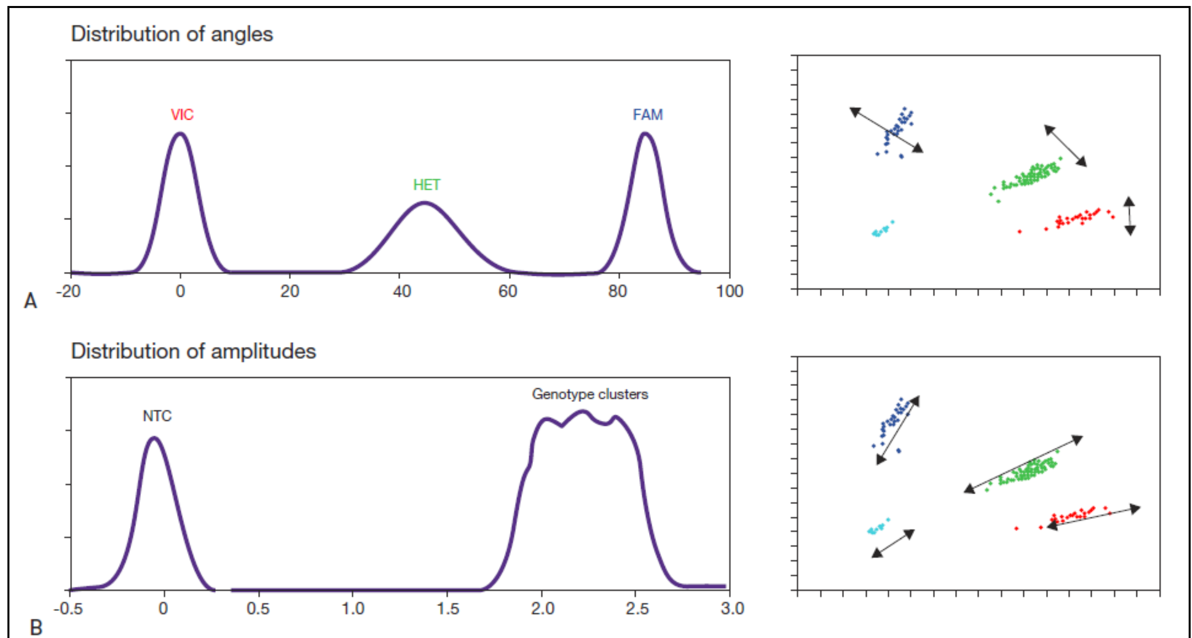


Figure 3-4: Distribution of angles (A) and amplitudes (B). The distribution of angles (A) shows three well-separated clusters. The distribution of amplitudes (B) shows three clusters with good signal intensity that is well separated from the NTC. [Taken from Life Technologies (2013)].

3.2. RATIONALE OF THE STUDY

Currently, there is an increased availability of validated and ready-to-use SNPs assays. Therefore, it has been an important factor to review these TaqMan® SNP Genotyping assays as a SNP genotyping system in detecting SNP markers.

In this study, two autosomal SNP markers (Sanchez et al. 2006a): rs1493232 (marker code 18) and rs1335873 (marker code 52) were selected to evaluate the TaqMan® SNP Genotyping assays (Applied Biosystems 2005). The TaqMan® Control Genomic DNA for rs1493232 is heterozygous with allele A and allele C, and homozygous for rs1335873 with only allele T. In the SNaPshot assay the alleles detected for rs1493232 are G and allele T. This is because the primers or probe used in TaqMan® SNP Genotyping assays are in forward orientation, whilst in SNaPshot assay (SBE reaction) they were in the reverse orientation (Sanchez et al. 2006a).

The analysis was carried out using the manufacturer's protocol (Applied Biosystems 2005, Applied Biosystems 2010a) and data were then analyzed using the TaqMan® Genotyper Software with a Quality value of 0.95 as a default (Life Technologies 2013).

3.3. RESULTS

3.3.1. Sensitivity and reproducibility of the TaqMan® SNP Genotyping assays

When employing any forensic tool, it is essential to carry out sensitivity and reproducibility studies. This is to determine the minimum concentration at which samples can reliably be typed. In this study, a series of dilutions (1 µl each) of (2.0, 1.0, 0.5, 0.25, 0.125, 0.063, 0.032, 0.016, 0.008, 0.004 and 0.002) ng/µl of Applied Biosystems TaqMan® Control Genomic DNA (Human; 10 ng/ µl) were used.

3.3.2 Preparation of Allelic Discrimination (AD) study

Several important steps and workflow were piloted before a number of assays were subjected into the analysis (Applied Biosystems 2005).

3.3.2.1. Preparation of the documents

Initially a few templates of AD documents were prepared in order to genotype the samples using the AB® 7500 Real-Time PCR System as shown in Appendix E. In creating a Pre-Read document, Allelic Discrimination assay was selected. New marker and detectors were appointed before the sample plate was set-up. Sample names and tasks (Unknown or No Template Control (NTC)) were labelled accordingly (Appendix E, Figure 11 to 14). As for AD Amplification (AQ) reaction template, assay type of Standard Curve (Absolute Quantitation) was selected. The same marker, detector, sample names with same positions (well number) and tasks were selected and labelled respectively, similar with the previous Pre-Read document (as shown in Appendix E, Figure 15 to 16). Finally, the AD Post-read document was also prepared, with an AD assay type, the similar marker, detector, sample names and tasks as above documents were selected (Figure 17 to 18).

3.3.2.2. Performing the run

The samples were analyzed using a few different conditions in each AD type of assays (as shown in Appendix E, Figure 19 to 21). In a Pre-Read analysis, the background of fluorescence used of each well of the of the subjected AD Pre-Read document was recorded. A Standard Amplification (AQ) run was performed only to amplify the samples using the selected PCR conditions, not use to quantify the PCR data. As for a post-read run, during the analysis, the Pre-Read fluorescence was subtracted from the post-Read fluorescence to interpret for pre-amplification background fluorescence in order to ensure the results were accurate.

3.3.2.3. Generating the results

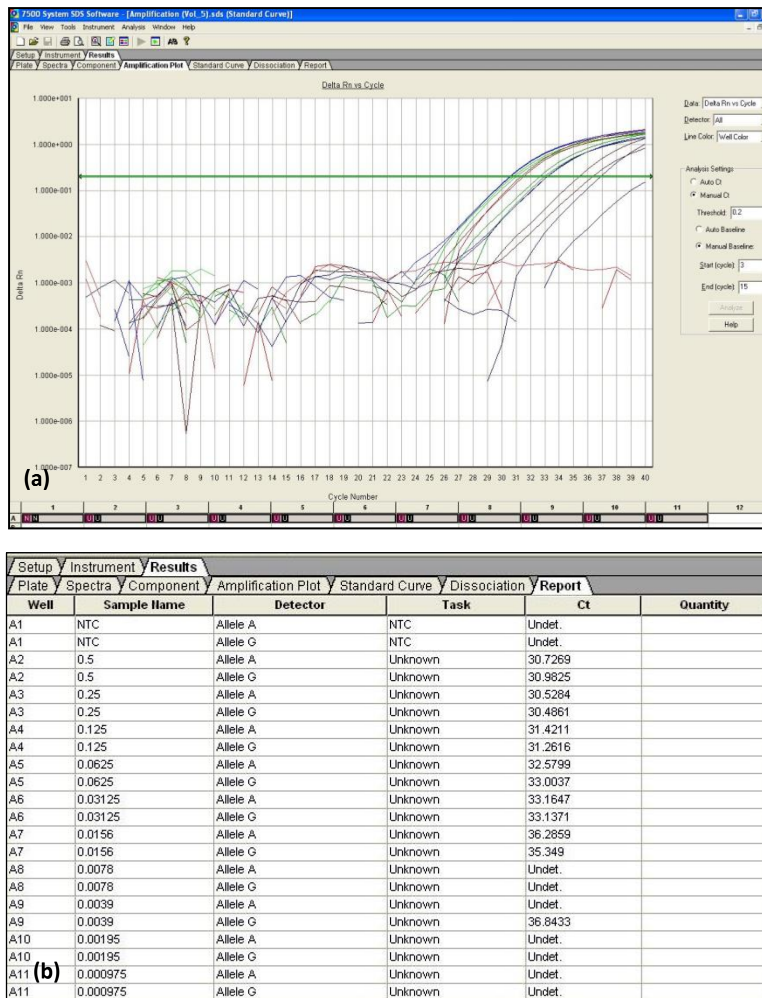
A few types of data were collected in the following analysis:

3.3.2.3.1. A Pre-Read data

As shown in Appendix E, Figure 22 and Figure 23, no alleles were observed and no report was generated. This is because in this type of analysis, the software only determines the fluorescence used in the assay.

3.3.2.3.2. *Amplification (AQ) data*

As for the AQ analysis, the amplification plot and real-time data (Figure 3-5 (a) and (b)) were observed and stored. The report produced in this stage also showed the targeted SNP marker. This analysis can be used to study the questionable allele-calls at the end of the whole AD process. However, no allele was observed in the Allelic Discrimination tab.



Well	Sample Name	Detector	Task	Ct	Quantity
A1	NTC	Allele A	NTC	Undet.	
A1	NTC	Allele G	NTC	Undet.	
A2	0.5	Allele A	Unknown	30.7269	
A2	0.5	Allele G	Unknown	30.9825	
A3	0.25	Allele A	Unknown	30.5284	
A3	0.25	Allele G	Unknown	30.4861	
A4	0.125	Allele A	Unknown	31.4211	
A4	0.125	Allele G	Unknown	31.2616	
A5	0.0625	Allele A	Unknown	32.5799	
A5	0.0625	Allele G	Unknown	33.0037	
A6	0.03125	Allele A	Unknown	33.1647	
A6	0.03125	Allele G	Unknown	33.1371	
A7	0.0156	Allele A	Unknown	36.2859	
A7	0.0156	Allele G	Unknown	35.349	
A8	0.0078	Allele A	Unknown	Undet.	
A8	0.0078	Allele G	Unknown	Undet.	
A9	0.0039	Allele A	Unknown	Undet.	
A9	0.0039	Allele G	Unknown	36.8433	
A10	0.00195	Allele A	Unknown	Undet.	
A10	0.00195	Allele G	Unknown	Undet.	
A11	0.000975	Allele A	Unknown	Undet.	
A11	0.000975	Allele G	Unknown	Undet.	

Figure 3-5: (a) An amplification plot of the AQ analysis, and (b) Report of the AQ analysis; that showed the targeted SNP marker and cycle threshold (C_T) of the samples.

3.3.2.3.3. Post-Read data

Data were analyzed using the 7500 SDS Software version 1.4, where the raw data were converted and expressed in terms of fluorescence signal versus filters. The SDS software plotted the results as shown in Figure 3-6, on a scatter plot of Allele X versus Allele Y. Initially, each well represented by X (Undetermined) on the plot. However, after analysis with Automatic/Manual Allele Calling settings, the clustering points were varied either along the horizontal axis (Allele X), vertical axis (Allele Y) or diagonal (Allele X and Allele Y). At this stage the genotype of the allele sample was determined (Applied Biosystems 2005). As shown below, after analysis three clusters were separated into 4 different positions, including the undetermined alleles. In addition, the report of the SNPs detected was also produced together with allele's position information (Figure 3-7).

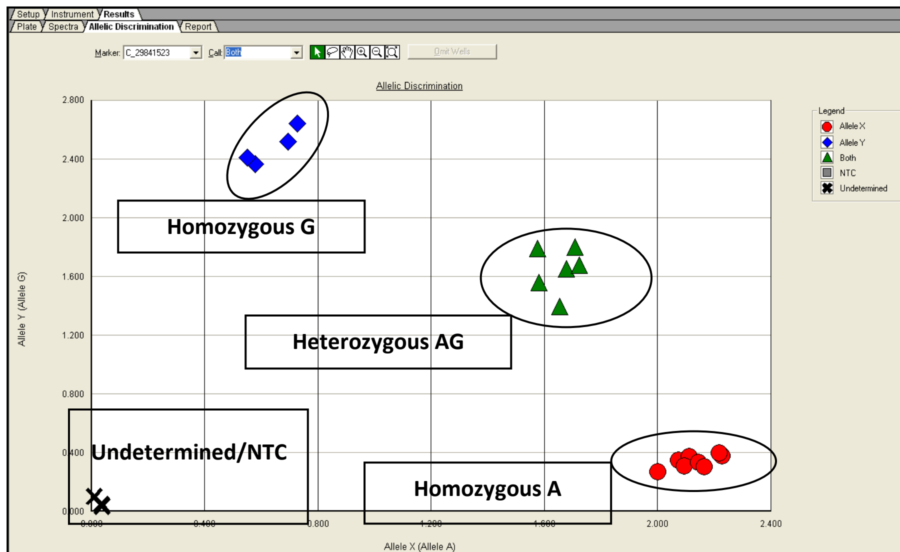


Figure 3-6: Detected alleles (homozygote A, homozygote G, heterozygote alleles AG and undetermined alleles) were plotted in Allelic Discrimination tab.

Setup | Instrument | Results
Plate | Spectra | Allelic Discrimination | Report

Well	Sample Name	Marker	Task	Pass.Ref	Allele X Delta Rn	Allele Y Delta Rn	Call	Quality(%)
E1	NTC	C_29841523	NTC	-2.96904e+004	0.0348403	0.0475586	Undetermined	100.00
E2	4	C_29841523	Unknown	-2.25517e+004	2.07359	0.346982	Allele A	100.00
E3	7	C_29841523	Unknown	-3.2176e+004	2.11224	0.372071	Allele A	100.00
E4	8	C_29841523	Unknown	1.19409e+004	2.00027	0.266697	Allele A	100.00
E5	11	C_29841523	Unknown	-3.00855e+004	2.09337	0.308994	Allele A	100.00
E6	17	C_29841523	Unknown	2.16256e+005	0.00987689	0.0991832	Undetermined	100.00
E7	18	C_29841523	Unknown	-2.66508e+004	1.65396	1.39382	Both	100.00
E8	32	C_29841523	Unknown	-2.47888e+004	1.67849	1.65305	Both	100.00
E9	33	C_29841523	Unknown	-2.34321e+004	2.22745	0.376844	Allele A	100.00
E10	53	C_29841523	Unknown	-2.29967e+004	0.694441	2.51986	Allele G	100.00
E11	64	C_29841523	Unknown	-1.77189e+004	1.57631	1.78988	Both	100.00
E12	74	C_29841523	Unknown	-1.498e+004	2.14333	0.332944	Allele A	100.00
F1	76	C_29841523	Unknown	-3.21915e+004	0.727149	2.64272	Allele G	100.00
F2	85	C_29841523	Unknown	-3.07811e+004	1.72336	1.67519	Both	100.00
F3	86	C_29841523	Unknown	-2.01533e+004	2.21627	0.399647	Allele A	100.00
F4	92	C_29841523	Unknown	-1.79002e+004	0.577303	2.36578	Allele G	100.00
F5	93	C_29841523	Unknown	-3.8163e+004	0.549974	2.40866	Allele G	100.00
F6	95	C_29841523	Unknown	-2.424e+004	1.58027	1.55994	Both	100.00
F7	99	C_29841523	Unknown	-3.17398e+004	2.16342	0.304411	Allele A	100.00
F8	100	C_29841523	Unknown	-2.40154e+004	1.70768	1.79983	Both	100.00
F9	NTC	C_29841523	NTC	-3.59342e+004	0.0363647	0.0344309	Undetermined	100.00

Figure 3-7: Report generated from the Post-Read analysis.

3.3.3. TaqMan® SNP Genotyping assays optimization

There are two main variables that are very important to assess in achieving the optimal conditions for TaqMan® SNP Genotyping assays: total PCR volume and template DNA concentration. For example, as shown in Figure 3-8, with higher amounts of template DNA, that ranged from 0.013 ng to 2 ng, the C_T (cycle threshold) was around 29 to 34. But, at lower DNA concentrations either the C_T was higher or DNA was not amplifiable at all.

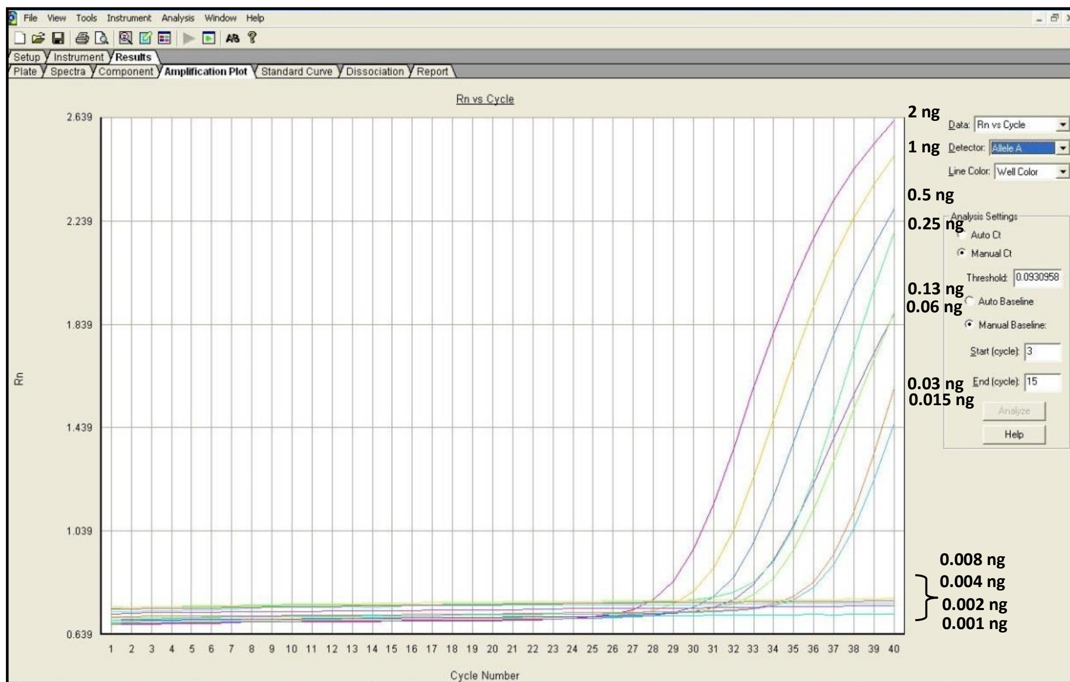
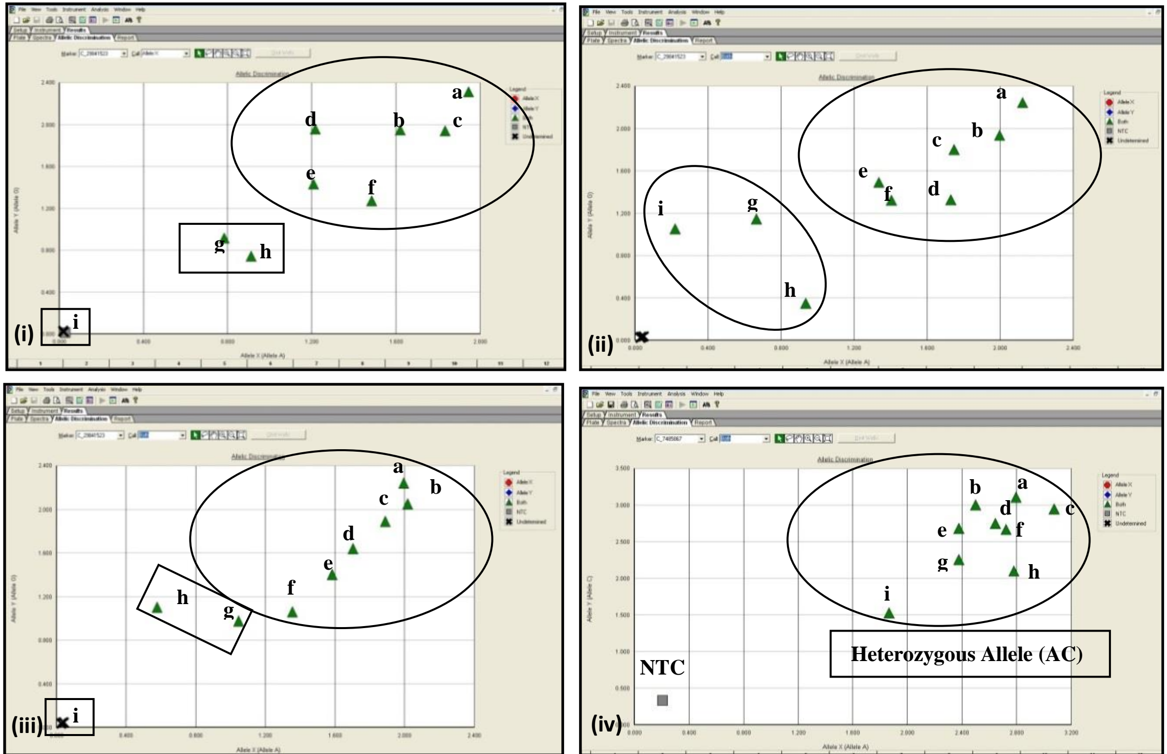


Figure 3-8: Effects of DNA concentrations on Rn versus cycles.

To test the effect of the PCR volume, variations in PCR reaction components were evaluated as described in *Chapter 2, section 2.7.1.2., Table 2-3*. Data presented in Figure 3-9 were generated by running the samples at low to high DNA concentrations and with various PCR volumes. The data plot and circles show samples with relatively high DNA concentration from 0.0625 ng to 2 ng, that are more consistent in developing the clusters for the relevant alleles compared to the samples with low DNA concentrations. The square boxes represent the unambiguous alleles and NTC.

With a total PCR volume of 25 µl results were reproducible for all series of DNA concentrations. All reactions with template DNA of 0.063 ng or more also gave reliable and reproducible results in all reaction volumes. In this study the final composition of the reaction mix and DNA concentrations chosen to be used were 5 µl of PCR volume and 0.125 ng of DNA concentration, due to its robustness as well as the reduced amount of reaction mix used in this system.

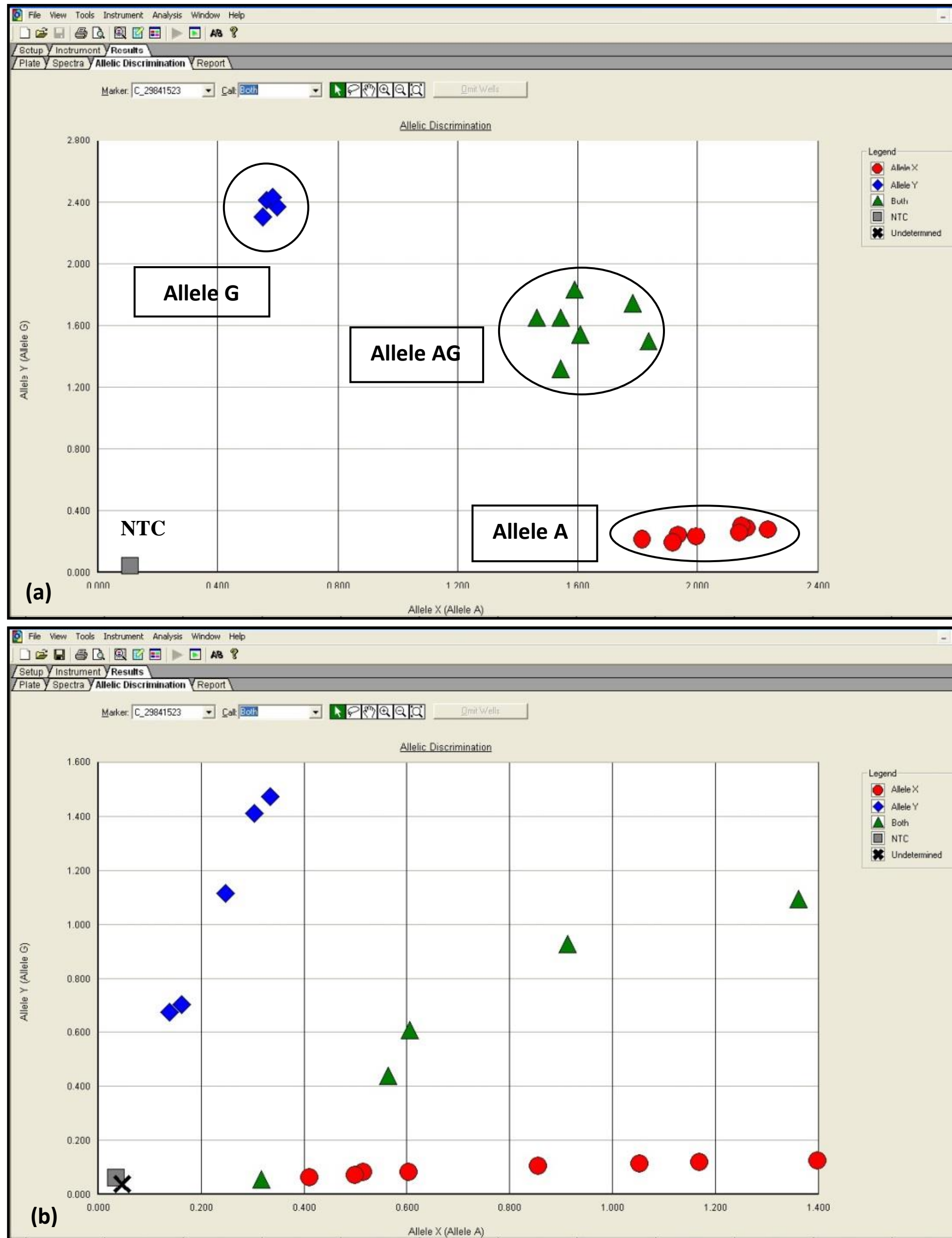
As part of the optimization, further study was carried out to observe clearly how robust this assay was especially in clustering the genotypes based on the optimal PCR but with two different DNA concentrations, that were 0.0078 ng and 0.125 ng.



Notes: *a, b, c, d, e, f, g, h and i represent (2, 1, 0.5, 0.25, 0.125, 0.063, 0.032, 0.016 and 0.008) ng, respectively. *x represent undetermined allele.

Figure 3-9: Genotype clusters and allele calls generated from the optimization of various PCR volumes and series of control DNA concentrations, (i) total PCR of 5 µl, (ii) 10 µl, (iii) 12.5 µl and (iv) 25 µl.

As shown in Figure 3-10, the data plot of population samples with (a) 0.125 ng DNA concentrations yields measurable results with good clustering of alleles. But, as for the samples with lower DNA the clusters became diffuse and it was difficult to identify the genotypes correctly and some were not detectable at all.



Note: *x represent undetermined allele.

Figure 3-10: Clusters of data generated from testing 19 population samples. All PCR reactions were in 5 µl, but with two different DNA concentrations; (a) 0.125 ng and (b) 0.0078 ng.

3.3.4. TaqMan® SNP Genotyping analysis on testing population samples

Thirty DNA samples from each ethnic group were genotyped under the optimal conditions: (i) a total of 5 µl PCR reaction volume, and (ii) 1 µl of DNA sample (0.125 ng). The analysis was carried out using specific TaqMan® MGB probes and primers of two chosen SNP markers: rs1493232 and rs1335873 (Sanchez et al. 2006a). Results showed the same alleles were clustered together and different types were clustered separately, for both SNP markers as shown in Figure 3-11(a) and (b).

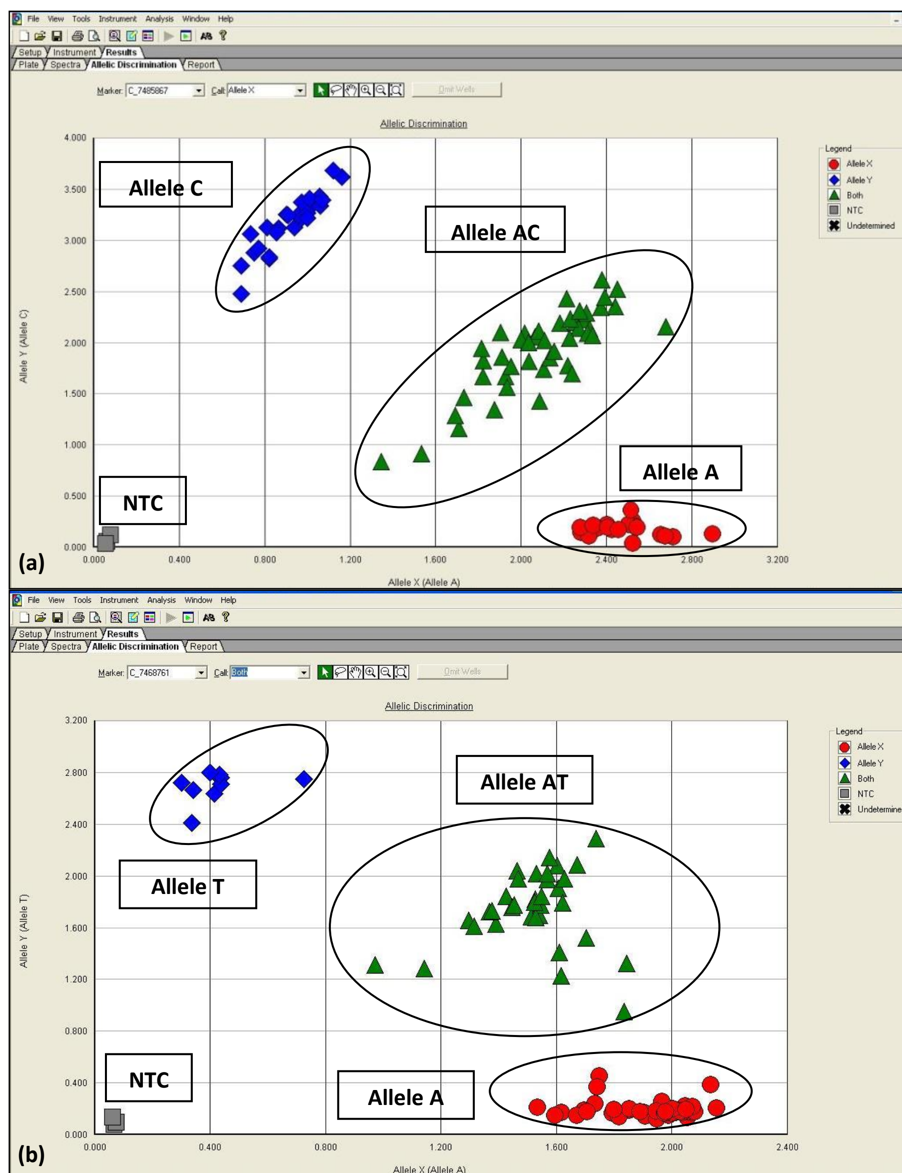


Figure 3-11: Clustered plots of 90 Malaysian DNA samples and 3 NTCs using the optimal conditions for 2 SNP markers: (a) rs1493232 (A/C), and (b) rs1335873 (A/T).

3.4. DISCUSSION

Real-time (RT) PCR methods for genotyping SNPs are becoming increasingly important in various fields of biological sciences (Gibson 2006). The advantage of this method is that only one simple reaction set up is required without any processing after that and it is suitable for both 96 or 384 sample reactions (Ragoussis 2006).

As for forensic casework, it is common to be faced with samples containing minute amounts of DNA and/or samples where the DNA is likely to be highly degraded. In some circumstances standard STRs are prone to fail in amplifying the DNA samples. To successfully resolve such cases an alternative marker, such as autosomal SNP typing can be used (Phillips et al. 2008a).

3.4.1. Allelic discrimination using TaqMan® SNP Genotyping assays and AB® 7500 Real- Time PCR System

In this study, TaqMan® SNP Genotyping assays were chose to be used because they provided the simplest workflow and a large number of validated SNP markers. Applied Biosystems Company, which is currently known as Life Technologies, offers the largest collection of ready-to-use human and mouse single nucleotide polymorphisms (SNP) assays available with TaqMan MGB probes. Human TaqMan® SNP Genotyping Assays collection consists of approximately 30,000 assays for SNPs in gene-coding regions, 160,000 gene centric assays and also more than 3 million genome-wide assays that include known disease mutations and SNPs in protein domains associated with drug binding regions (Applied Biosystems, 2006).

Secondly, the use of the TaqMan probes provides the specificity of PCR amplification. There are two types of TaqMan probes available: TaqMan probes with TAMRA™ dye as quencher and FAM™, TET™ or VIC® as 5' label dye; TaqMan MGB (minor groove-binding) probes with nonfluorescent quencher (NFQ) and FAM™, TET™, NED™ or VIC® as 5' label dye are recommended when the probes exceed 30 nucleotides. The non-fluorescent quencher allows the instrument to measure the reporter dye contributions more precisely. The MGB probes have greater differences in melting temperature

values between matched and mismatched probes, which allow more accurate allelic discrimination ((Applied Biosystems 2006, Pakstis et al. 2007).

As stated in the product protocol, the reaction needs only three components: genomic DNA template, SNP Genotyping Assay and PCR Master Mix (TaqMan® Universal PCR Master Mix). The standard protocol of PCR amplification is valid for each assay: 1 cycle for 10 min at 95 °C to activate the AmpliTaq Gold enzyme and 40 cycles of 15 sec at 92 °C, 1 min at 60 °C. These simple steps reduce the chance of contamination, sample mix-up and sample loss, and the need to optimize each reaction separately (Applied Biosystems 2006).

Besides that, the interpretation of data is easy because results/alleles will be displayed in cluster forms, and report will be presented as a table format. This TaqMan SNP Genotyping Assay is read at the PCR endpoint rather than in real time incontrasted to TaqMan assays designed for either absolute or relative quantitation (as in gene expression studies). Genotype calls for individual samples are made by plotting the normalized intensity of the reporter dyes in each sample well on a Cartesian plot (also known as a scatter or cluster plot). A clustering algorithm in the data analysis software assigns individual sample data to a particular genotype cluster. SNP Genotyping Assays produce three clusters biallelic SNPs (De La Vega et al. 2005a).

The instrument, such as AB® 7500 can meet all throughput needs and budgets. The 7500 offers variable excitation and 5-colour detection via FAM/SYBR Green, VIC/JOE, NED/TAMRA/Cy3, ROX/Texas Red, and Cy5. Detection is superior in the red of the spectrum compared to the 7300 system. The format is Peltier-based thermal cycling systems allowing the use of both 96-well plates and 0.2 ml tubes (Gibson 2006). In addition, following PCR amplification, an assay endpoint can be read on any Applied Biosystems Real-Time PCR System.

3.4.2. Detection level of the TaqMan® SNP Genotyping assays on few parameters

To develop an efficient multiplex PCR, careful planning, several tests in primer designing and balancing reactions are required (Butler 2011). In this study, variations in

PCR volumes reaction and DNA concentration for each reaction were assessed to achieve the optimal conditions for the TaqMan® SNP assays. According to the manufacturer's protocol, a total of 25 µl PCR volumes per well in a 96-well plate was recommended, in which 13.75 µl PCR mix is added with 11.25 µl of DNA (1-20 ng with DNase-free water).

However, from the study's evaluation, the reduced volume of PCR-TaqMan reactions with 0.125 ng genomic DNA and 5 µl total PCR reaction mix gave reproducible signals that were assigned to clear genotypes called. Signals obtained on reactions with less than 0.125 ng were not clearly plotted or clustered, and thus no accurate genotyping was possible with these low concentrations using this assay (as shown in Figure 3-10 (b)).

The samples with relatively high DNA concentrations were easily scored, but with samples with lower DNA concentrations, such as 0.01 ng and 0.1 ng, clusters became diffuse making the alleles scoring difficult.

Zubakov et al. (2010), stated that the high sensitivity of TaqMan® MicroRNA Assays makes them potentially valuable tools for future forensic applications. He and his co-workers analyzed 20 ng to 2 pg of RNA with TaqMan® MicroRNA Assays using real-time PCR. They found that all four markers were detected in the target body fluid using the lowest amount of total RNA (2 pg) for cDNA synthesis (Zubakov et al. 2010).

3.4.3. Limitation of TaqMan® SNP assays

From this study's observations, which are based on the sensitivity and reproducibility data, TaqMan® SNP genotyping assays can be used as a SNP genotyping platform. A large number of SNP assays have been validated and pre-designed. However, these assays will just involve one reaction and one marker per run.

Vega et.al. (2005)(De La Vega et al. 2005a) agreed that the TaqMan® SNP assay offers sensitive, robust and fast genotyping using real-time methods and would be useful in detecting single reaction or one targeted SNP marker at one time. It is also only

involves a single reaction which might be useful and suitable for the selection of markers for genetic studies and simple forensic casework (De La Vega et al. 2005a, Kidd et al. 2006, Pakstis et al. 2007).

As for forensic purposes, a medium throughput is required for paternity testing or criminal casework, but a high throughput would be needed to implement criminal DNA databases. An important factor that is a limiting step for forensic genetics in all these technologies is the amount of DNA required per genotype. The main drawback of this TaqMan technology is the limited multiplexing capability (Sobrino et al. 2005b).

Kidd and co-workers (Kidd et al. 2006) stated that they used TaqMan for the screening procedures because they were screening markers individually and did not have to develop or optimize the assays. Multiplex assays will be essential in any actual forensic application but TaqMan is not capable of being multiplexed. Therefore, a SNP panel will require a different typing method.

PART TWO:

SNAPSHOT ASSAYS

3.1. INTRODUCTION

Multiplex PCR is an extension of the standard PCR protocol in which multiple loci are amplified simultaneously in order to save time, improve throughput, and reduce the total cost of reagents and amount of DNA (Rachlin et al. 2005a).

As for this study, 52 SNP markers (Sanchez et al. 2006a) were chosen to be evaluated on the Malaysian population and casework samples. This selection was made based on the features of these markers, including the number of SNP markers involved, the amplicon size of each SNP marker and its potential in human identification. In addition, these markers have also been reported to be polymorphic in European, Asian and African populations (Sanchez et al. 2006a, Musgrave-Brown et al. 2007).

Gill calculated that when using binary polymorphisms between 50 and 80 loci are required to match the discrimination levels of 16 STRs (Gill 2001). Dixon has suggested that to develop new SNP multiplexes for forensic identification purposes, there are a number of factors that have to be considered. One of them is to have low molecular weight genomic targets, specifically lower than current conventional STR systems with the amplicon size preferably less than 150 bp (Dixon et al. 2005).

According to the original paper (Sanchez et al. 2006a), these 52 SNP markers were amplified in one PCR reaction followed by detection of SNPs with two single-base

extension (SBE) reactions by Capillary Electrophoresis (CE) as shown in Figure 3-12. The two SBE multiplexes comprised the first 23 primers (marker codes 1 to 23) and another 29 primers (marker codes 24 to 54). The PCR product sizes ranged from 59 bp to 115 bp, whilst the SBE products size ranged from 16 bp to 92 bp. The original system was tested using the SNaPshot® Multiplex System to detect the SNP genotypes on three genetic analyzers: ABI 310 PRISM®, ABI 3100 PRISM® or ABI 3100 PRISM® Avant, all with 36 cm capillary arrays and performance optimized polymer, POP-4™.

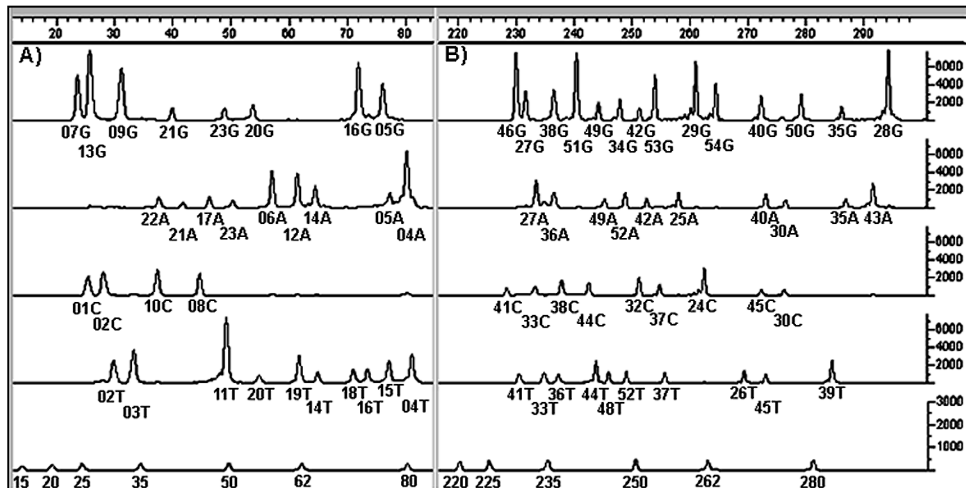


Figure 3-12: A 52 SNP marker assay with two sets of SBE reactions; A) 23 markers, and B) 29 markers. [Taken from the original study (Sanchez et al. 2006a)]

In addition, Sanchez et al. (2006a) also stated that the sensitivity of the ABI 310 PRISM® was only 20% of the sensitivity of the ABI 3100 PRISM®, and therefore most of their studies were carried out using higher version systems, i.e., the ABI 3100 and ABI 3100-Avant PRISM® Genetic Analyzers.

There are many other studies have been carried out using these markers for several purposes, such as population study, routine casework and paternity testing. Schwark et al. (2012b) presented the use of 50-plex SNP assay of SNPforID 52plex on several routine cases from their laboratory. Samples were subjected to two different multiplex reactions, 21-plex and 29-plex assays (which they termed Auto 1 and Auto 2). They reported that 29-plex (Auto 2) was more sensitive than the 21-plex (Auto 1) when using ABI 310 genetic

analyzers. With input DNA less than 500 pg, fewer successfully amplified SNPs were found in the 21-plex, with only 5 SNPs successfully analyzed, but half of the 29 SNPs were still detectable with DNA input as low as 25 pg. They further suggested to use the 29-plex PCR first and then, depending on the outcome, subject the sample material to the less-sensitive 21-plex in routine work (Schwark et al. 2012b).

Bulbul et al. (2009) have performed an internal validation of 29-plex (Auto 2) using ABI 310 and demonstrated all 29 samples were typed successfully with between 0.5 ng and 20 ng DNA input. They also found that even 0.2 ng gave results although allele drop out was observed (Bulbul et al. 2009).

3.2. RATIONALE OF THE STUDY

Although SNPs will not replace STRs in the near future as the primary source of information used in criminal investigations, it is advantageous to assess and validate DNA typing platforms for reliable, accurate and high throughput multiplex SNP typing to be used along with PCR-STRs. Therefore, the objectives of this study are as the following:

3.2.1. Sensitivity and reproducibility evaluation of the SNaPshot assay(s)

The sensitivity and reproducibility of the SNaPshot assay were carried out as mentioned in *Chapter 2, section 2.6.1.*, for multiplex reactions, and for single reactions, 0.5 ng and 1 ng of TaqMan Control DNA and M7-7 (Malay individual sample) were used as the controls throughout the study. To ensure the appropriate SBE size, correct SNP markers and alleles exhibited, single SNaPshot reactions were performed (as shown in Figure 3-13) for all 52 markers using both controls, TaqMan DNA and M7-7 (as described in *Chapter 2, sections 2.7.2.3 to 2.7.2.8.1*)

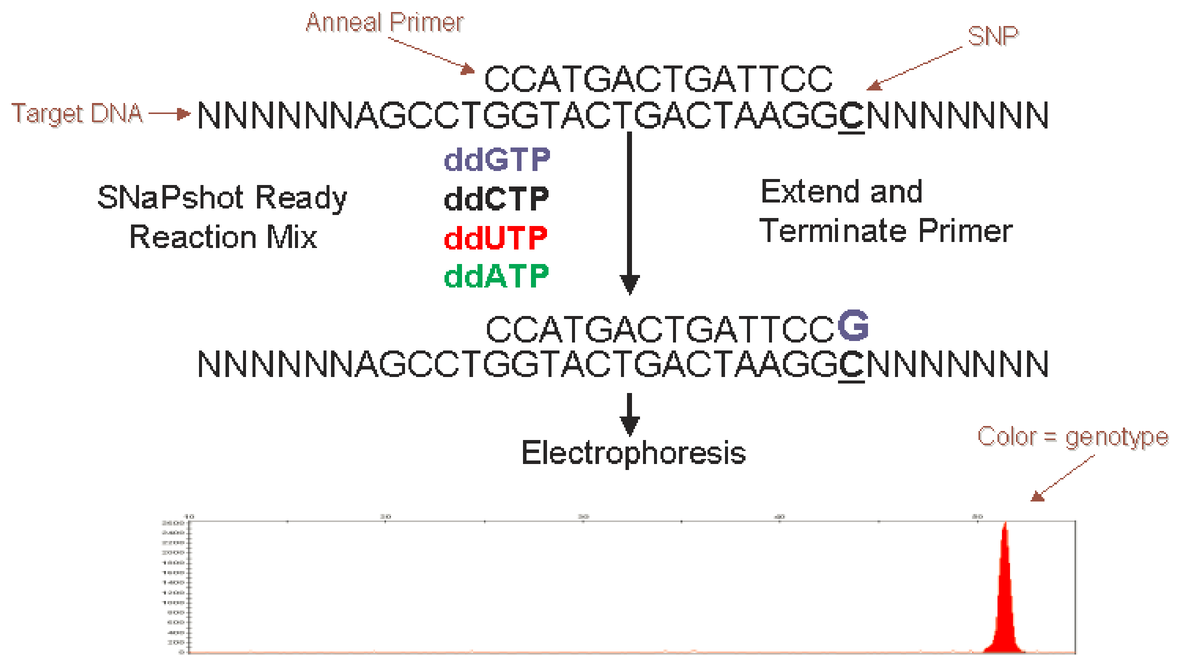


Figure 3-13: Single nucleotide primer extension [Taken from Applied Biosystems, 2010b].

3.2.2. Developing a series of multiplex assays to be used in the study

A few series of multiplex assays was developed to observe the effect of primer-primer interactions, dyes and peak height exhibited and also to test the PCR/SBE systems used in the study.

3.2.3. Optimization and validation of the selected SNaPshot multiplex assays

Based on experimental work, a series of multiplex assays were chosen to be used. These multiplexes comprised four sets of PCR/SBE 13-plex assays, which designated as 13_{st}, 13_{nd}, 13_{rd} and 13_{th}. Optimization and validation studies had been carried out for these assays to observe their stability and reproducibility level.

3.2.4. SNP genotyping and characterization of 30 unrelated testing individuals

Genotyping analysis of 30 samples from each Malaysian ethnic group (Malays, Chinese and Indians) was carried out using these 13-plex assays. These samples had already been analyzed using the TaqMan® genotyping method for comparison in the detection of SNP alleles.

SNP analysis had also been carried out on 9947A (Positive control from Applied Biosystems) using protocol developed in this study and the result was sent to Dr. Christopher Phillips from SNPforID Consortium. We had passed the assessment that shows concordance result with the Consortium. This also confirmed the dyes labelled and SNPs detected were all correct with the peak-ratio criteria as suggested by Sanchez et. al. 2006a.

3.3. RESULTS

3.3.1. Single SNaPshot reaction

Initially, several primer concentrations and combinations were tested for PCR-SBE reactions as described in *Chapter 2, section 2.7.2.1*. Each of the PCR-SBE reactions used 1 µl of control DNA (0.5 ng). The allele peak generated from PCR/SBE primers combination (Table 3-1) was shown as in Figure 3-14.

Table 3-1: Table shows the combination of primers concentration used in the study.

Stage	Combination of the primer concentrations (µM)					
PCR	0.05 (a)	0.05 (b)	0.1 (c)	0.1 (d)	0.2 (e)	0.2 (f)
SBE	0.05	0.1	0.05	0.1	0.05	0.1

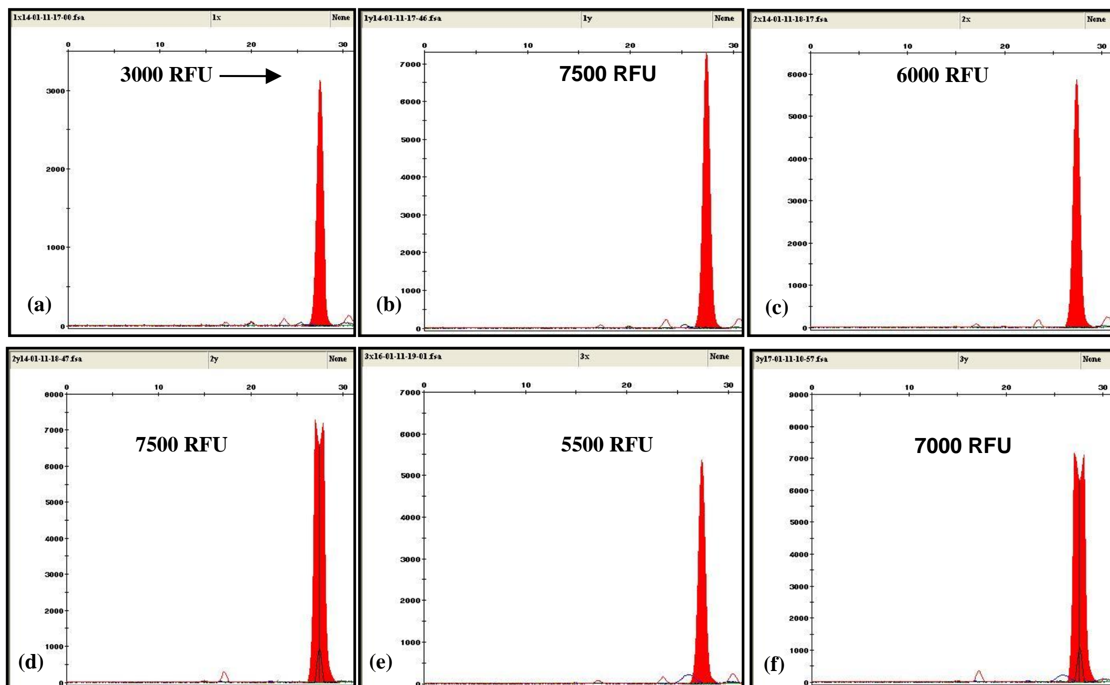


Figure 3-14: Amplicons generated from the single reaction assay with the effects of primer concentrations.

The data demonstrated the effects of primer concentrations. For example, as shown in Table 3-1, although the primer concentrations are as low as 0.05 (for both, PCR and SBE reactions), the allele peak generated was 3000 RFU, which is considered adequate for SNP detection.

The size, dye and allele type for each 52 SNP markers of both controls were recorded and used in evaluating each SNP marker in the multiplex assays.

3.3.2. Series of multiplexes developed throughout the study

In order to multiplex the markers several trials that involved several combinations of markers were carried out as listed below:

3.3.2.1. First trial of developing 3-plex, 5-plex and 10-plex assays

Initially multiplexes with 3-, 5- and 10-plexes of markers were conducted as shown in Table 3-2. Single reaction amplification was also performed for each individual marker, in order to confirm the correct target was detected.

Table 3-2: The SNP markers with primer concentrations (µM) used in the multiplex assays (*first trial*).

Multiplexes with SNP code markers, primers concentrations (µM) and SBE size (nt)											
3-plex				***5-plex***				***10-plex***			
Marker	**PCR**	**SBE**	**SBE size**	**Marker**	**PCR**	**SBE**	**SBE size**	**Marker**	**PCR**	**SBE**	**SBE size**
3	*0.3*	*0.2*	29	**1**	*0.3*	*0.2*	18	**2**	*0.1*	*0.2*	24
***10**	*0.4*	*0.3*	60(34)	**3**	*0.3*	*0.2*	29	**3**	*0.15*	*0.2*	29
***18**	*0.3*	*0.3*	53(66)	***4**	*0.3*	*0.3*	36(78)	***4**	*0.15*	*0.2*	36(78)
				***18**	*0.3*	*0.3*	53(66)	***6**	*0.15*	*0.3*	44(54)
				***10**	*0.3*	*0.3*	60(34)	***7**	*0.2*	*0.2*	48(18)
								***10**	*0.2*	*0.2*	60(34)
								***12**	*0.2*	*0.3*	29(58)
								***17**	*0.15*	*0.3*	49(42)
								***18**	*0.1*	*0.2*	53(66)
								***20**	*0.25*	*0.2*	60(50)

Note: * indicates modified SBE primers length. Number in brackets indicates the original SBE amplicon size (Sanchez et al. 2006a).

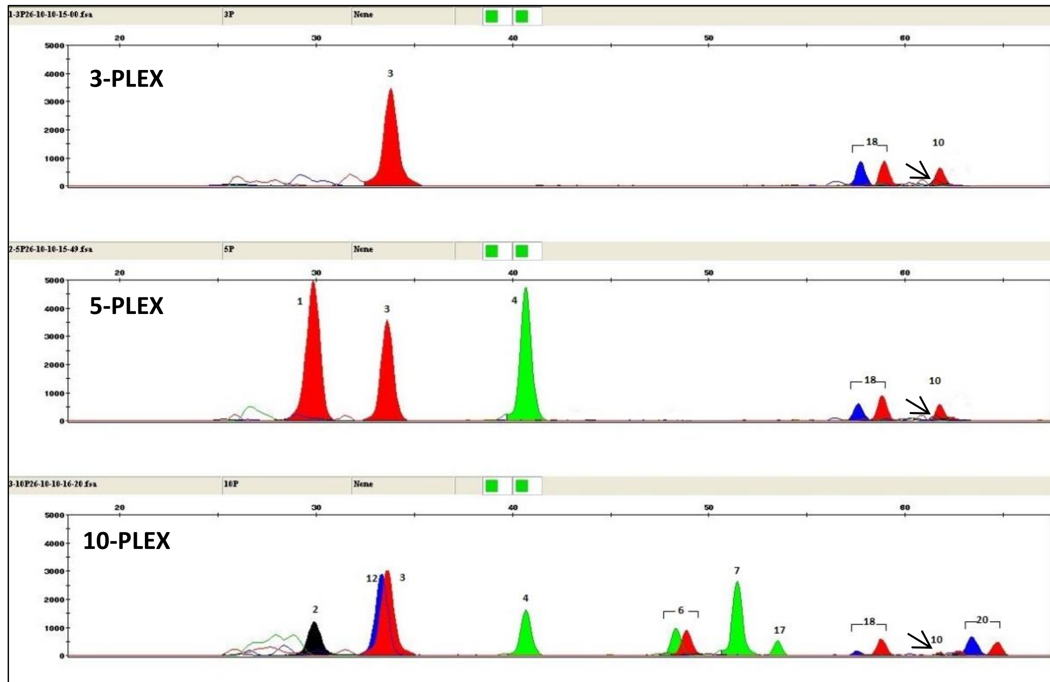


Figure 3-15: Three multiplex assays; 3-plex, 5-plex and 10-plex of TaqMan control DNA (0.5 ng). Arrows show the drop-out allele C (for code marker 10).

As shown above in Figure 3-15, there is a drop-out allele for SNP marker 10, which should be heterozygous CT, but only allele T was detected in all multiplexes. This phenomenon could be due to lower primer concentrations for marker 10 and also preferential amplification of the markers. The difference in SBE product size for code marker 10 (CT) and 18 (CT) is 7 nucleotides (nt) length.

With the 10-plex assay a few markers were chosen that had similar SBE product size to observe if there are any significant primer-primer reactions that will lead to 'drop-in' or 'drop-out'. From Table 3-2, marker 3 and 12 (29 nt), and marker 10 and 18 (60 nt) have the same product size. After the analysis, as can be seen from Figure 3-15, homozygous alleles 3 and 12 are clear, but as for heterozygous alleles 10 and 20, only allele GT (marker 20) can be clearly observed, whilst the heterozygous alleles CT dropped-out.

3.3.2.2. Second attempt at developing 3-plex, 5-plex and 10-plex assays with new SBE sizes.

Table 3-3: Multiplex assays consisting of SNP markers with different SBE size and primer concentrations.

Multiplexes with SNP code markers, primers concentrations (µM) and SBE size (nt)											
3-plex2				***5-plex2***				***10-plex2***			
Marker	**PCR**	**SBE**	**SBE size**	**Marker**	**PCR**	**SBE**	**SBE size**	**Marker**	**PCR**	**SBE**	**SBE size**
3	*0.3*	*0.2*	29	**1**	*0.3*	*0.2*	18	**1**	*0.1*	*0.07*	18
10	*0.3*	*0.3*	34	**3**	*0.3*	*0.2*	29	**2**	*0.1*	*0.05*	24
***18**	*0.4*	*0.3*	45(66)	**4**	*0.3*	*0.3*	47	**3**	*0.1*	*0.06*	29
				***18**	*0.3*	*0.3*	45(66)	***4**	*0.15*	*0.15*	47(78)
				10	*0.3*	*0.3*	34	***5**	*0.15*	*0.3*	70(74)
								6	*0.1*	*0.2*	54
								***7**	*0.1*	*0.05*	20(18)
								***8**	*0.1*	*0.15*	40(42)
								***9**	*0.1*	*0.05*	30(29)
								10	*0.2*	*0.15*	34

Note: * indicates modified SBE primer length. Number in brackets indicates the original SBE amplicon size (Sanchez et al. 2006a).

The results obtained from the analysis of the multiplex assays (Table 3-3 and Figure 3-16), shows that the smaller the size of SBE product, the clearer the alleles (as shown in 3-plex2 and 5-plex2). Whilst, for the 10-plex2, if the bigger SBE product size was incorporated, larger differences in SBE length (nt) were needed to produce good and clear genotypes. However, in this experiment, the SNP marker 2 was not detected, in both, single and 10-plex2 reactions. New primers (with same sequences) for PCR and SBE of marker 2 were re-ordered.

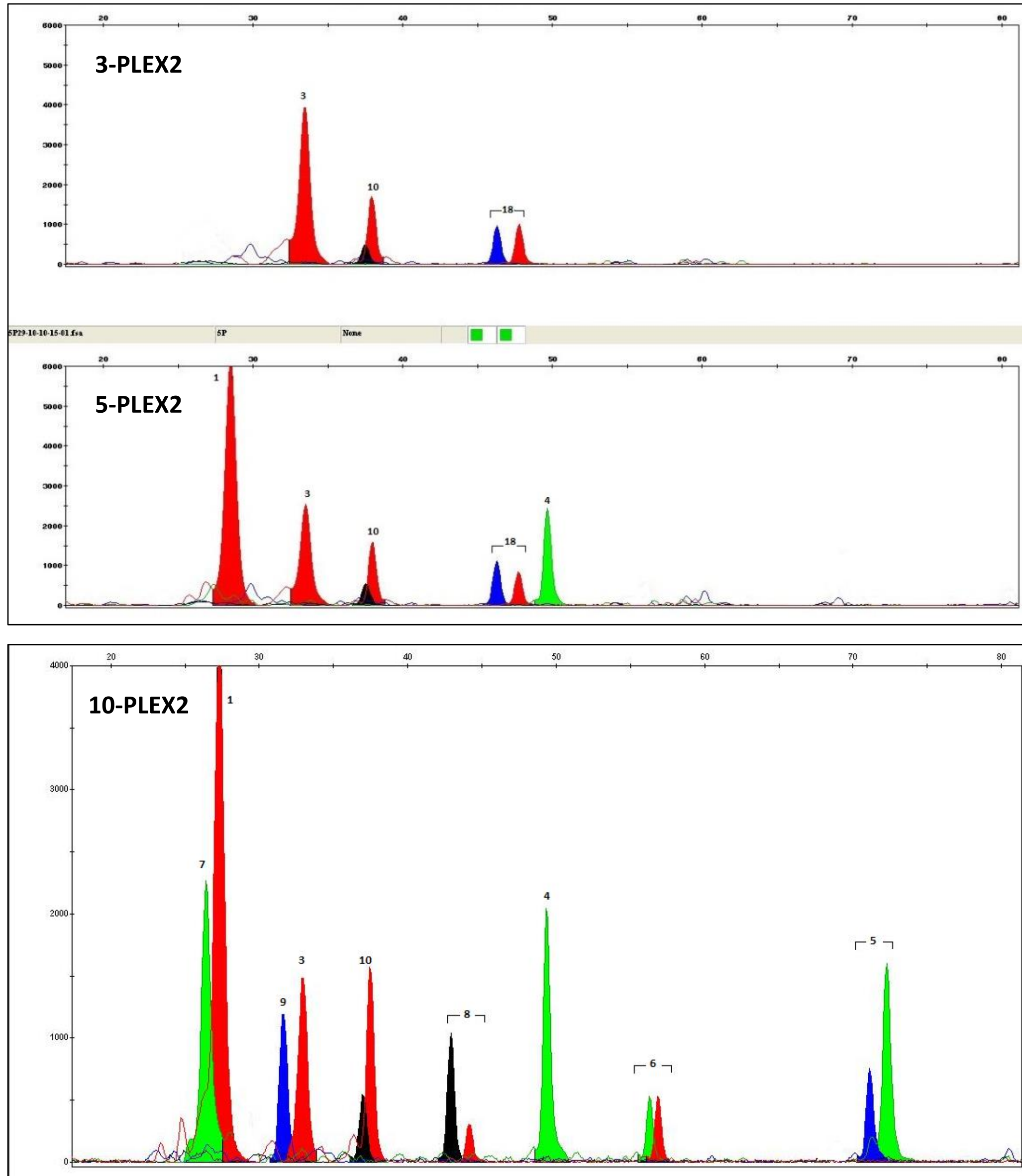


Figure 3-16: Three different multiplex assays for the second trial in multiplexing the markers; 3-plex2, 5-plex2 and 10-plex2 of TaqMan control DNA (0.5 ng).

3.3.2.3. *First trial in developing four sets of 13-plex assays*

Based on the results obtained from the trials (*section 3.3.1 to 3.3.2.2*), the original SNP markers (Sanchez et al. 2006a) were divided into four sets of 13-plex assays, 1st-plex, 2nd-plex, 3rd-plex and 4th-plex (Table 3-4). The markers were re-arranged in accordance with the SBE product size, to ensure clear separation of alleles in all multiplex assays. The first two 13-plex assays, 1st-plex and 2nd-plex, comprised SNP markers of code number 1 to 26 (Sanchez et al. 2006a), whilst the other two 13-plex assays comprised SNP markers of code number 27 to 54. Both controls, TaqMan Control DNA and M7-7 (each, 0.5 ng and 1 ng) were analyzed with all 13-plex assays. For better visualization and interpretation of the sample, all 13-plex assays were presented as a whole profile (52 autosomal SNP markers) with 4 panels per individual.

Table 3-4: The first construction of four sets 13-plex PCR/SBE assays.

Four 13-plex assays with SNP code markers and primers concentrations (µM)											
1st			***2nd***			***3rd***			***4th***		
Marker	**PCR**	**SBE**	**Marker**	**PCR**	**SBE**	**Marker**	**PCR**	**SBE**	**Marker**	**PCR**	**SBE**
1	*0.1*	*0.07*	**7**	*0.15*	*0.05*	**41**	*0.15*	*0.05*	**46**	*0.1*	*0.1*
2	*0.15*	*0.05*	**13**	*0.1*	*0.06*	**33**	*0.1*	*0.06*	**27**	*0.1*	*0.07*
3	*0.1*	*0.06*	**9**	*0.1*	*0.05*	**36**	*0.15*	*0.07*	**38**	*0.15*	*0.08*
21	*0.1*	*0.05*	**10**	*0.15*	*0.15*	**44**	*0.1*	*0.08*	**51**	*0.1*	*0.1*
22	*0.15*	*0.1*	**8**	*0.2*	*0.15*	**48**	*0.15*	*0.07*	**49**	*0.15*	*0.1*
17	*0.15*	*0.15*	**11**	*0.15*	*0.1*	**34**	*0.15*	*0.09*	**52**	*0.15*	*0.1*
23	*0.1*	*0.15*	**4**	*0.1*	*0.15*	**32**	*0.1*	*0.1*	**42**	*0.1*	*0.1*
20	*0.15*	*0.1*	**25**	*0.1*	*0.15*	**37**	*0.15*	*0.15*	**53**	*0.15*	*0.1*
6	*0.15*	*0.2*	**24**	*0.1*	*0.15*	**29**	*0.1*	*0.15*	**54**	*0.2*	*0.1*
12	*0.1*	*0.2*	**19**	*0.2*	*0.1*	**40**	*0.1*	*0.1*	**45**	*0.15*	*0.1*
14	*0.1*	*0.15*	**26**	*0.1*	*0.1*	**30**	*0.15*	*0.1*	**50**	*0.1*	*0.15*
16	*0.15*	*0.2*	**18**	*0.1*	*0.2*	**39**	*0.1*	*0.15*	**35**	*0.1*	*0.15*
5	*0.15*	*0.3*	**15**	*0.1*	*0.3*	**43**	*0.15*	*0.15*	**28**	*0.15*	*0.1*

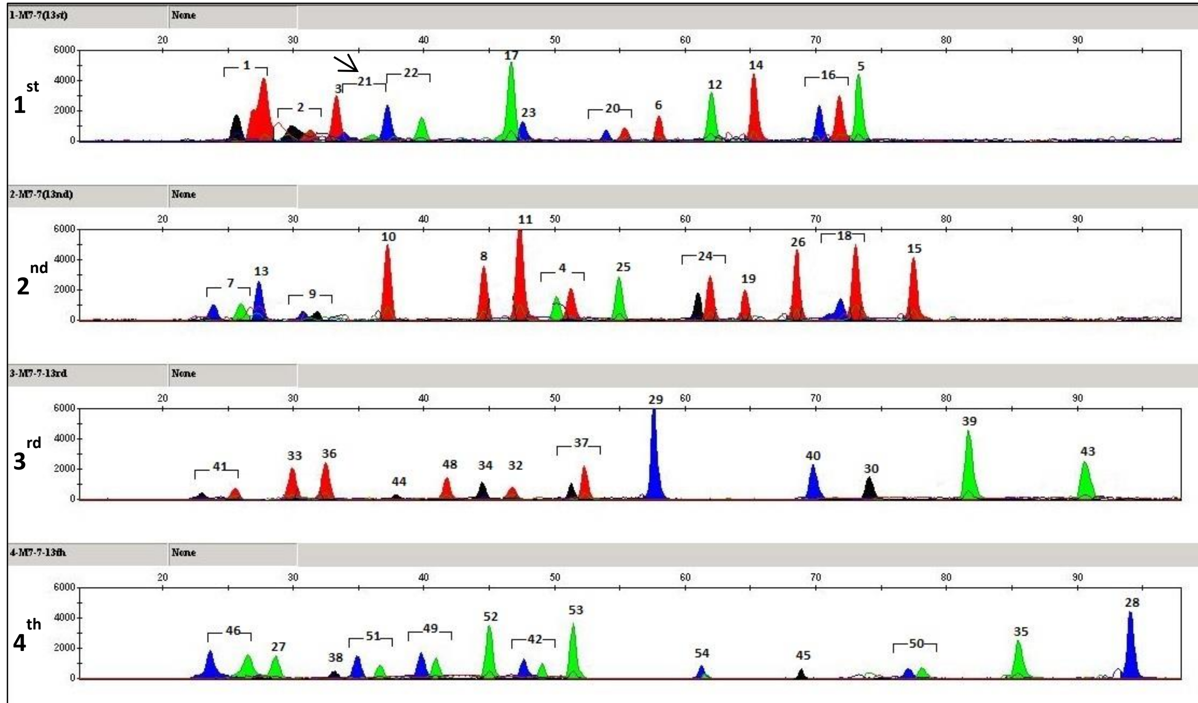


Figure 3-17: An example of the first 13-plex assays (1st-plex, 2nd-plex, 3rd-plex and 4th-plex) of M7-7 DNA sample (0.5 ng) developed in this study.

In Figure 3-17, these multiplex assays (as combinations in Table 3-4) showed the alleles were separated well from each other except for the heterozygote alleles of SNP marker 21 (shown with an arrow). Also, the alleles in 1st and 2nd assays were quite close to each other compared to the other two 13-plex assays, 3rd and 4th. This is because the SBE size for marker 1 to 26 for these assays ranged from 16 nt to 80 nt. As for the other two 13-plex assays, the SBE size in these developed assays ranged from 16 nt to 92 nt.

In addition, the dyes exhibited in the 2nd-plex and 3rd-plex had more red dyes, compared to the other two 13-plex assays, whilst the 4th-plex had more green dyes exhibited. This presentation of SNP profile could potentially lead to misinterpretation of data.

Therefore, another experiment was carried out using a different arrangement of the SNP markers, in order to balance the size and dyes produced in the analysis. The organization of the new four sets of 13-plex assays was presented as in Table 3-5.

3.3.2.4. Second trial in developing four sets of 13-plex assays

Table 3-5: Second trial of 13-plex assays developed in the study.

Four 13-plex assays with SNP code markers and primers concentrations (µM)											
1^{st2}			*2^{nd2}*			*3^{rd2}*			*4^{th2}*		
Marker	**PCR**	**SBE**	**Marker**	**PCR**	**SBE**	**Marker**	**PCR**	**SBE**	**Marker**	**PCR**	**SBE**
1	*0.1*	*0.07*	**7**	*0.15*	*0.08*	**41**	*0.15*	*0.1*	**46**	*0.1*	*0.1*
2	*0.15*	*0.1*	**13**	*0.1*	*0.06*	**27**	*0.1*	*0.1*	**33**	*0.1*	*0.1*
3	*0.1*	*0.08*	**9**	*0.1*	*0.15*	**36**	*0.15*	*0.08*	**38**	*0.15*	*0.15*
22	*0.15*	*0.1*	**10**	*0.15*	*0.15*	**21**	*0.1*	*0.15*	**51**	*0.1*	*0.1*
34	*0.15*	*0.1*	**8**	*0.2*	*0.15*	**44**	*0.1*	*0.15*	**48**	*0.15*	*0.1*
11	*0.15*	*0.1*	**17**	*0.15*	*0.15*	**49**	*0.15*	*0.1*	**52**	*0.15*	*0.1*
23	*0.1*	*0.15*	**4**	*0.1*	*0.15*	**32**	*0.15*	*0.15*	**42**	*0.1*	*0.1*
20	*0.15*	*0.2*	**25**	*0.1*	*0.15*	**53**	*0.15*	*0.1*	**37**	*0.15*	*0.15*
6	*0.15*	*0.25*	**24**	*0.1*	*0.15*	**29**	*0.1*	*0.1*	**54**	*0.2*	*0.15*
12	*0.1*	*0.2*	**19**	*0.2*	*0.15*	**40**	*0.1*	*0.1*	**45**	*0.15*	*0.2*
14	*0.1*	*0.15*	**26**	*0.1*	*0.1*	**30**	*0.15*	*0.15*	**50**	*0.15*	*0.15*
16	*0.15*	*0.2*	**18**	*0.1*	*0.2*	**39**	*0.1*	*0.15*	**35**	*0.1*	*0.15*
5	*0.15*	*0.3*	**15**	*0.1*	*0.3*	**43**	*0.15*	*0.15*	**28**	*0.15*	*0.1*

In this second arrangement a few markers; 11, 17, 21, 27, 33, 34, 37, 48, 49 and 53, were swapped into different multiplex assays (Table 3-5). As illustrated in Figure 3-18, the 13-plex assays profile of M7-7 individual showed clearer separation of alleles. However, it is noticeable there are alleles which overlapped with other alleles, such as marker 7, and also some of the markers have preferential amplification over other markers, for example marker 20 and marker 6 (shown with the arrows). Therefore, another change in markers positions was carried out to reduce the preferential amplification among the targeted loci. The final designated four sets of 13-plex assays: *13_{st}, 13_{nd}, 13_{rd} and 13_{th}* were introduced as in Table 3-6.

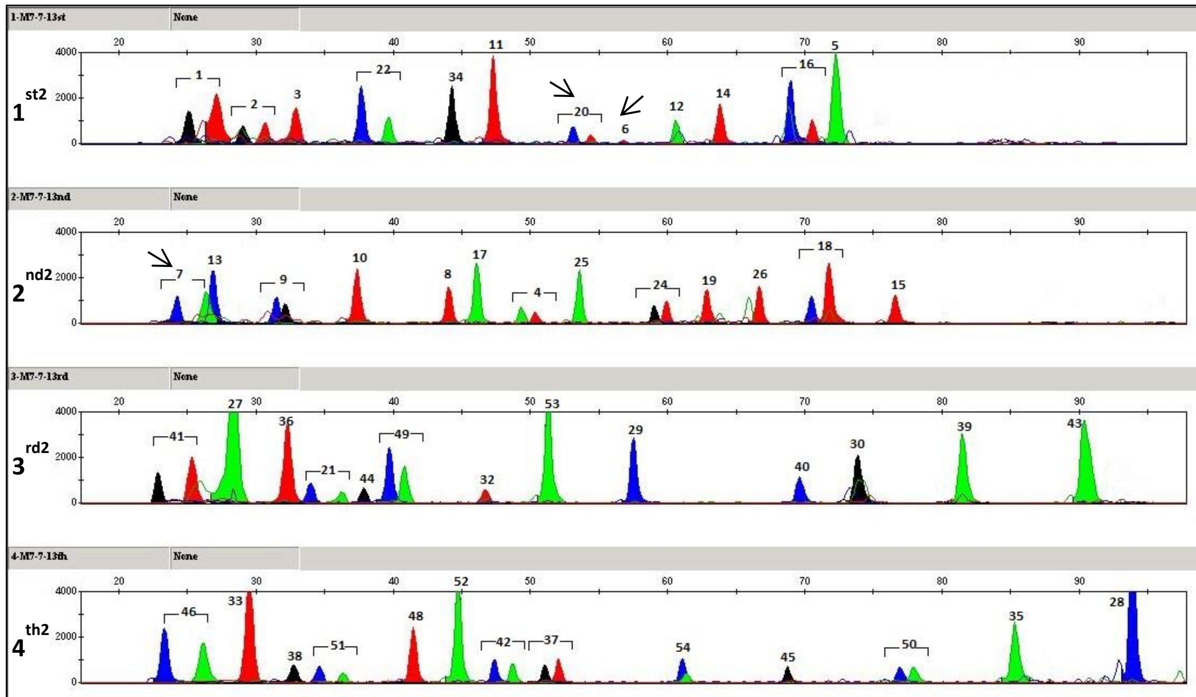


Figure 3-18: An example of the second 13-plex assays (1^{st2}-plex, 2^{nd2}-plex, 3^{rd2}-plex and 4^{th2}-plex) of M7-7 DNA sample (0.5 ng) developed in this study.

3.3.3. Development, optimization and validation of the 13-plex PCR/SNaPshot reactions; 13_{st}, 13_{nd}, 13_{rd} and 13_{th}.

Finally, after all criteria were evaluated, such as primer-primer interactions, correct allele dyes obtained and peak height threshold, four sets of 13-plex PCR and SNaPshot reactions were developed in this study to carry out the SNP analysis on the Malaysian population. Table 3-6 showed the modification of the new 13-plex assays from the original multiplex assays. The original SBE size (Sanchez et al. 2006a) was indicated in the brackets for the SNP marker that labelled with * symbol.

The sensitivity study of these 13-plex assays was carried out using a series of control DNA dilutions. All analysis was performed as described earlier in *Chapter 2, sections 2.7.2.3 to 2.7.2.8.1*. Results obtained from the analysis (as shown in Figures 3-19 to 3-22) demonstrate 13-plex assays (*13_{st}, 13_{nd}, 13_{rd} and 13_{th}*) of TaqMan® Control Genomic DNA serial dilutions of (1.0, 0.5, 0.125 and 0.03) ng. Figure 3-23, shows an example of a marker

bin (13_{nd} assays) that can be created using the GeneMapper version 3.2 and used to characterize the SNP alleles.

The peak thresholds were set to a minimum of 120 RFUs (blue colour), 60 RFUs (green colour) and 30 RFUs (yellow, red and orange colours). A maximum peak height ratio of 3:1 was accepted as a heterozygote, and a minimum peak of 5:1 was accepted as a homozygote (Sanchez et al. 2006a).

In all series of DNA concentrations, complete 52 autosomal SNP profiles were successfully generated under all tested conditions with more than 30 ng of template. In Figure 3-19, the profile is nearly complete, but allele C for marker 32 dropped-out. This shows that the thermal conditions, PCR components, SNaPshot reactions, the combination of PCR and SBE primers concentrations were established for an optimal performance of these multiplex assays for DNA input higher than 30 pg.

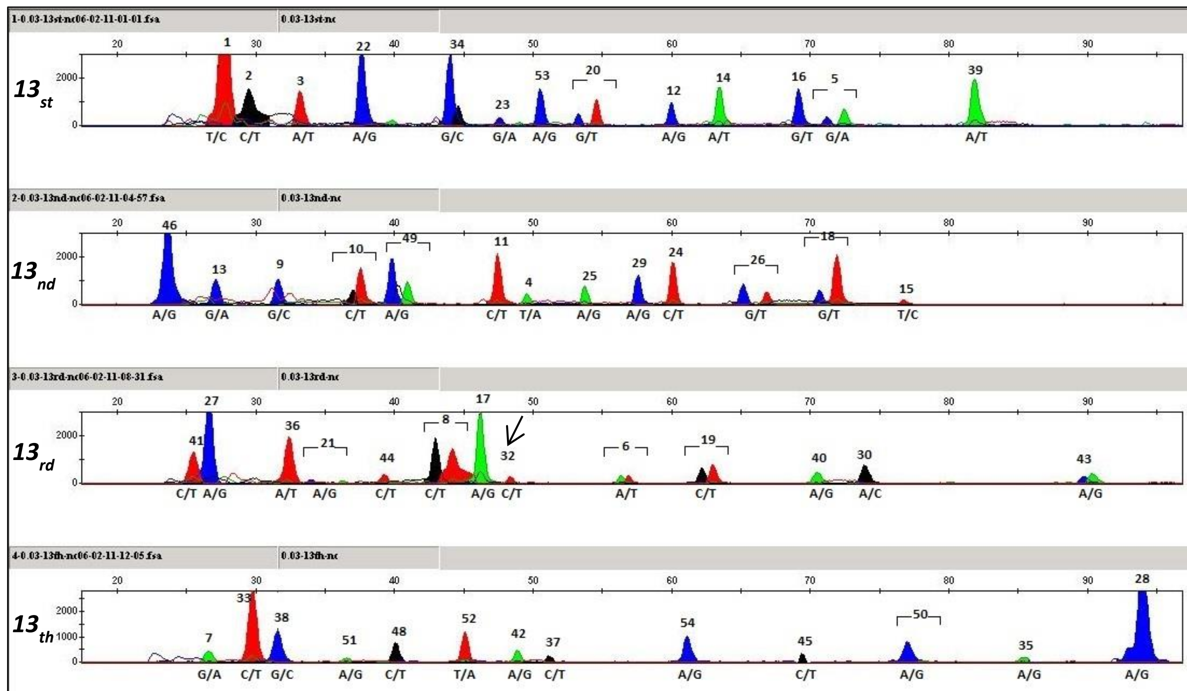


Figure 3-19: Human control DNA with 1 µl of serial dilution (0.03 ng). The arrow shows an allele C that dropped-out.

Table 3-6: Four sets of 13-plex assay reactions developed in this study.

Sanchez *et al.*, 2006 (Original SBE Multiplexes)			New Developed Assays in this study			PCR		Single base extension (SBE)	
SBE Multiplex	Marker Code	SNP (rs#)	Designated multiplex assays	Marker Code	SNP (rs#)	Final primer concentration (μM)	Amplicon size (bp)	Final primer concentration (μM)	Amplicon size (bp)
Auto-1 (23 SBE)	1	rs1490413	***13_{st}***	1	rs1490413	0.15	68	0.07	18
	2	rs876724		2	rs876724	0.1	83	0.1	24
	3	rs1357617		3	rs1357617	0.15	90	0.08	29
	4	rs2046361		22	rs733164*	0.1	68	0.1	36 (34)
	5	rs717302		34	rs1979255	0.15	86	0.15	40
	6	rs1029047		23	rs826472*	0.2	85	0.2	45 (46)
	7	rs917118		53	rs1028528	0.1	113	0.1	48
	8	rs763869		20	rs1031825	0.15	98	0.2	50
	9	rs1015250		12	rs2107612	0.15	93	0.3	58
	10	rs735155		14	rs1454361	0.1	73	0.15	62
	11	rs901398		16	rs729172*	0.1	60	0.2	68 (70)
	12	rs2107612		5	rs717302*	0.15	86	0.3	70 (74)
	13	rs1886510		39	rs354439	0.1	93	0.2	80
	14	rs1454361	***13_{nd}***	46	rs1360288	0.1	103	0.07	16
	15	rs2016276		13	rs1886510*	0.1	86	0.06	26 (25)
	16	rs729172		9	rs1015250*	0.1	95	0.15	30 (29)
	17	rs740910		10	rs735155	0.15	100	0.15	34
	18	rs1493232		49	rs1005533	0.1	107	0.1	36
	19	rs719366		11	rs901398*	0.15	70	0.1	44 (46)
	20	rs1031825		4	rs2046361*	0.3	79	0.15	47 (78)
	21	rs722098		25	rs873196	0.1	63	0.15	52
	22	rs733164		29	rs1024116	0.1	76	0.1	56
	23	rs826472		24	rs2831700	0.1	62	0.15	56
Auto-2 (23 SBE)	24	rs2831700		26	rs1382387	0.13	69	0.1	64
	25	rs873196		18	rs1493232*	0.1	59	0.2	68 (66)
	26	rs1382387		15	rs2016276	0.15	90	0.3	80
	27	rs2111980	***13_{rd}***	41	rs737681	0.1	96	0.07	16
	28	rs2056277		27	rs2111980	0.05	72	0.05	23
	29	rs1024116		36	rs2076848	0.1	89	0.08	27
	30	rs727811		21	rs722098*	0.15	81	0.15	32 (38)
	32	rs1413212		44	rs914165	0.1	100	0.15	32
	33	rs938283		8	rs763869*	0.15	100	0.2	40 (42)
	34	rs1979255		17	rs740910	0.1	87	0.15	42
	35	rs1463729		32	rs1413212	0.3	84	0.3	44
	36	rs2076848		6	rs1029047	0.25	100	0.4	54
	37	rs1355366		19	rs719366*	0.2	105	0.2	60 (58)
	38	rs907100		40	rs2040411	0.1	94	0.1	68
	39	rs354439		30	rs727811	0.15	78	0.15	72
	40	rs2040411		43	rs251934	0.15	98	0.2	88
	41	rs737681	***13_{th}***	7	rs917118*	0.1	87	0.15	20 (18)
	42	rs2830795		33	rs938283	0.08	85	0.1	22
	43	rs251934		38	rs907100	0.15	91	0.15	27
	44	rs914165		51	rs891700*	0.1	109	0.1	31 (32)
	45	rs10495407		48	rs964681	0.15	106	0.15	36
	46	rs1360288		52	rs1335873	0.15	110	0.1	40
	48	rs964681		42	rs2830795	0.1	97	0.15	44
	49	rs1005533		37	rs1355366	0.15	90	0.15	48
	50	rs8037429		54	rs1528460	0.15	115	0.15	60
	51	rs891700		45	rs10495407	0.4	102	0.2	68
	52	rs1335873		50	rs8037429	0.2	108	0.15	76
	53	rs1028528		35	rs1463729	0.1	87	0.15	84
	54	rs1528460		28	rs2056277	0.07	73	0.1	92

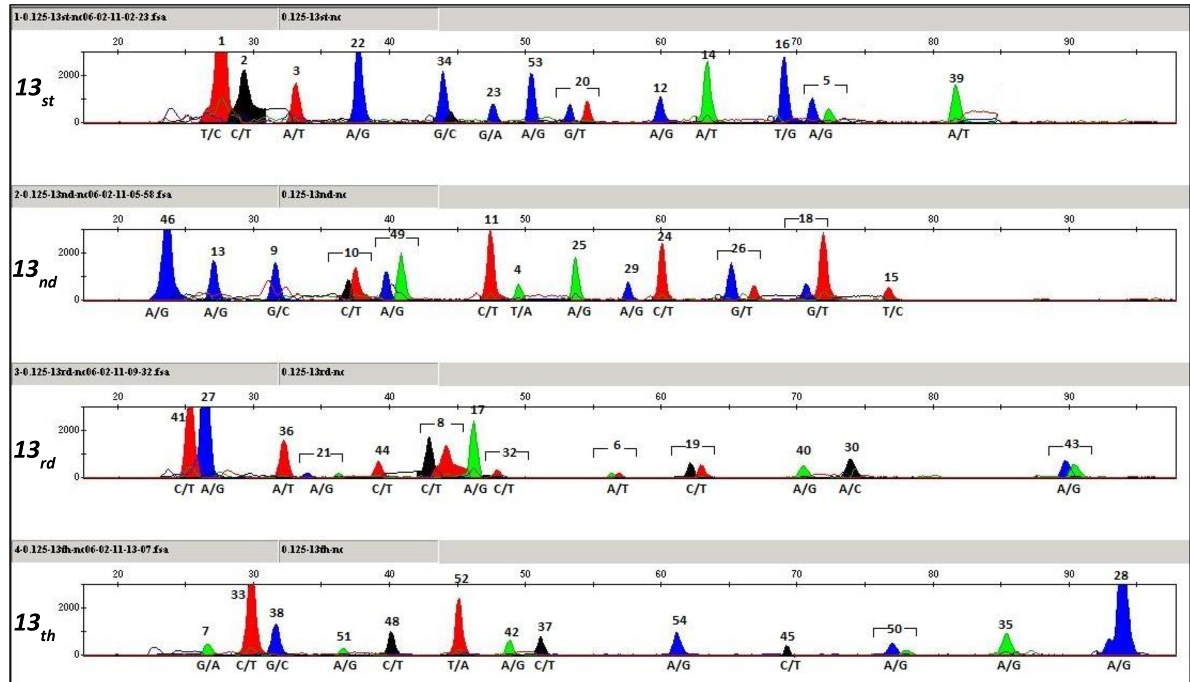


Figure 3-20: Full SNP profile of human control DNA with 1 µl of serial dilution (0.125 ng).

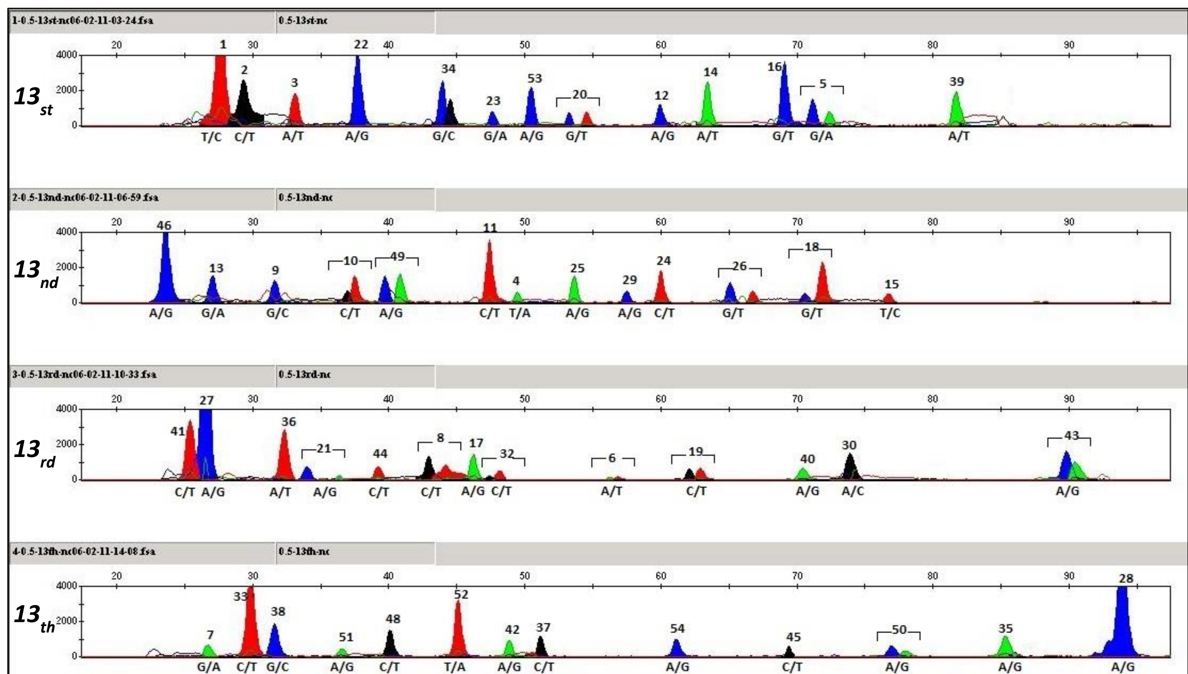


Figure 3-21: Full SNP profile of human control DNA with 1 µl of serial dilution (0.5 ng).

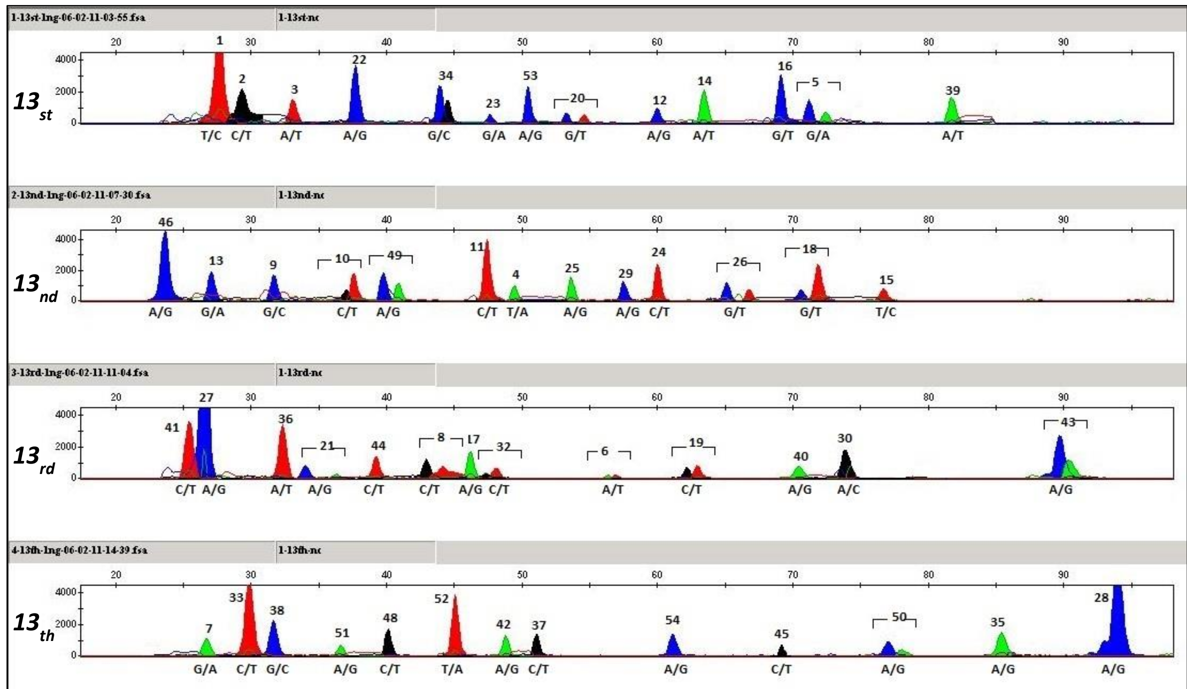


Figure 3-22: Full SNP profile of human control DNA with 1 µl of serial dilution (1.0 ng).

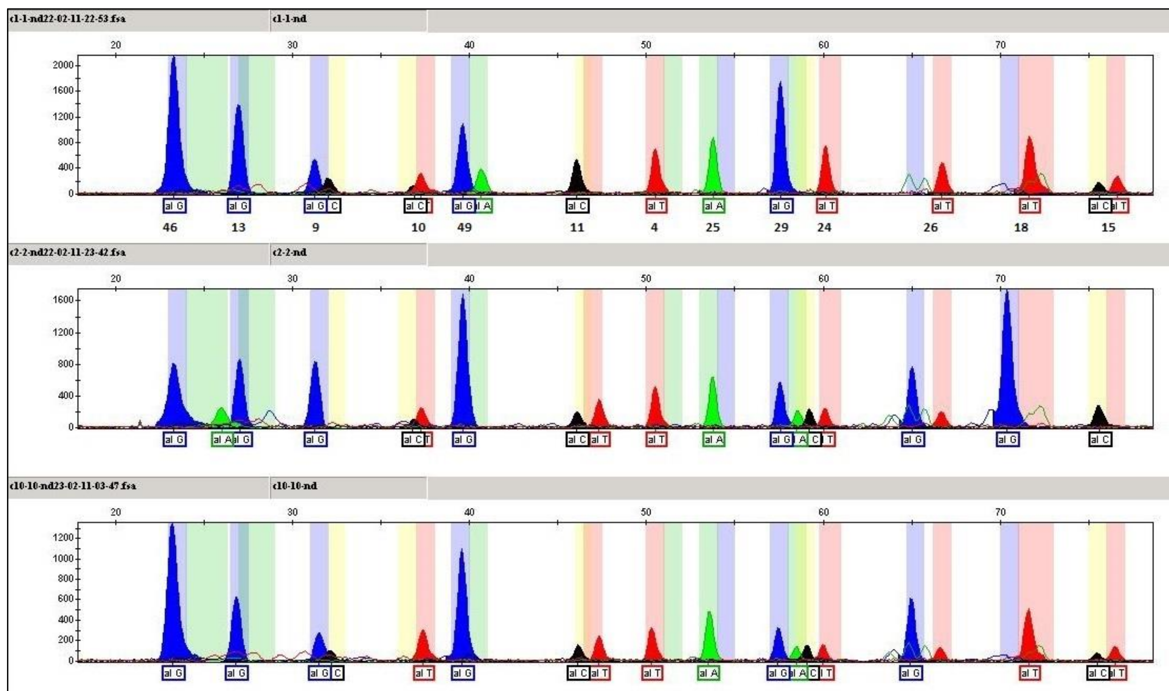


Figure 3-23: Three population samples (1 ng each) analyzed using the 13_{nd} bin created using the GeneMapper v 3.2.

3.3.4. Evaluation of 13-plex SNaPshot assays with testing population samples

Using these assays a further study was carried out on the population. A wide range of DNA concentrations of the testing population samples (0.04 ng to 569 ng) was used to assess the reproducibility of the 13-plex SNaPshot assays. The samples tested included 10 samples that were classed as undetermined by the Quantifler™ Human DNA Quantification kit (Applied Biosystems, USA). These samples when subjected to SNaPshot analysis typically resulted in partial profiles and only 3 samples produced full SNP profiles. Representatives of full and partial profiles are shown as in Figure 3-24 and 3-25, respectively. These samples were re-extracted using Organic Solvent (phenol/chloroform/isoamyl alcohol) method, quantified with Quantifler™ Human DNA Quantification kit (Applied Biosystems, USA) and then, re-genotyped to collect the SNP data.

Overall, full complete SNP profiles were successfully generated from all 90 population samples, at input DNA as low as 0.04 ng and as high as 70 ng, and no drop-out or unusual peaks were observed. However, peak imbalances and allele drop-out were observed in samples at lower DNA amounts that less than 0.04 ng or higher than 70 ng (Figure 3-16). In addition, the background noise was quite prominent in samples with higher template levels. Two examples of SNP profiles produced from the samples with input DNA approximately 70 ng and 569 ng are shown in Figures 3-26 and 3-27, respectively.

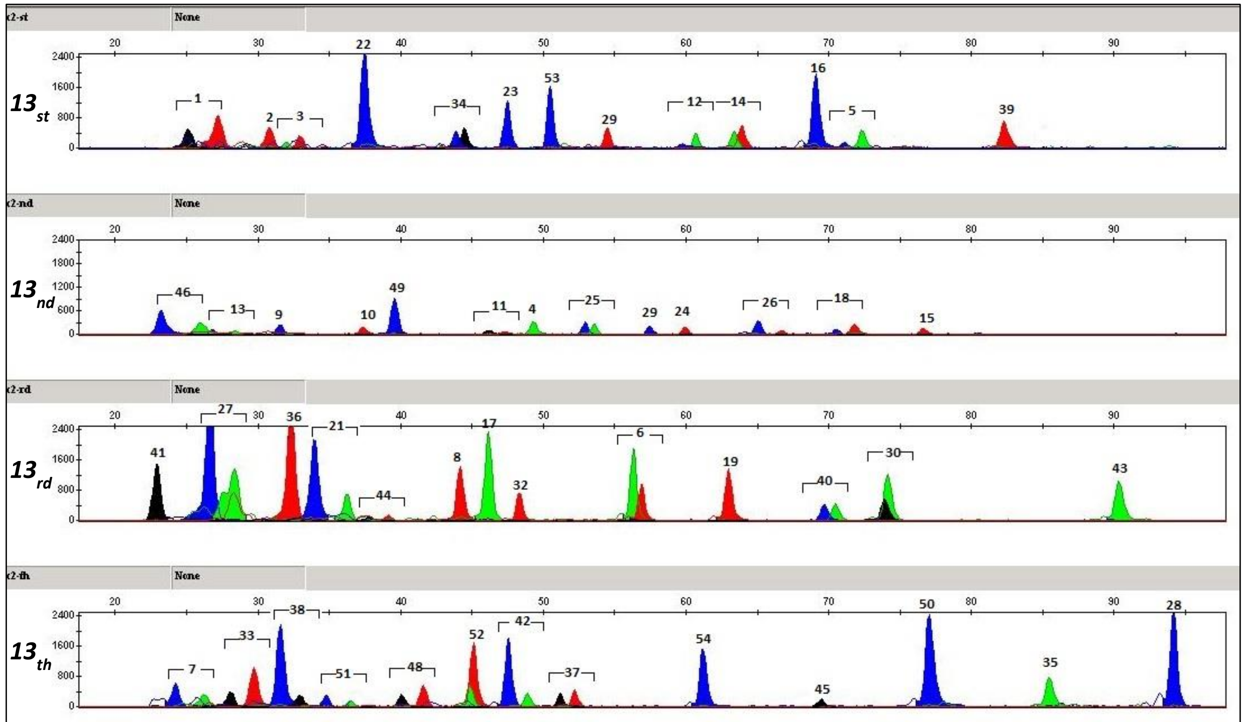


Figure 3-24: Full SNP profile of an individual with 0.04 ng DNA input.

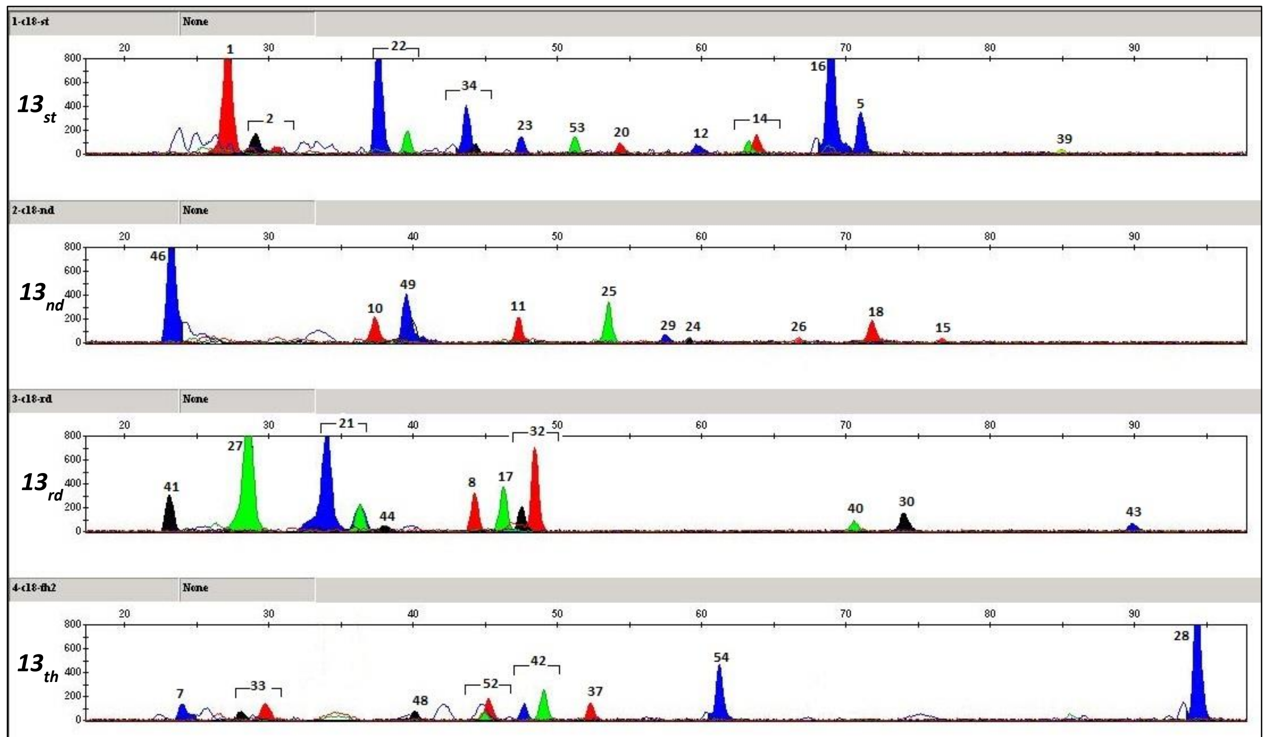


Figure 3-25: Partial SNP profile generated from undetermined DNA template.

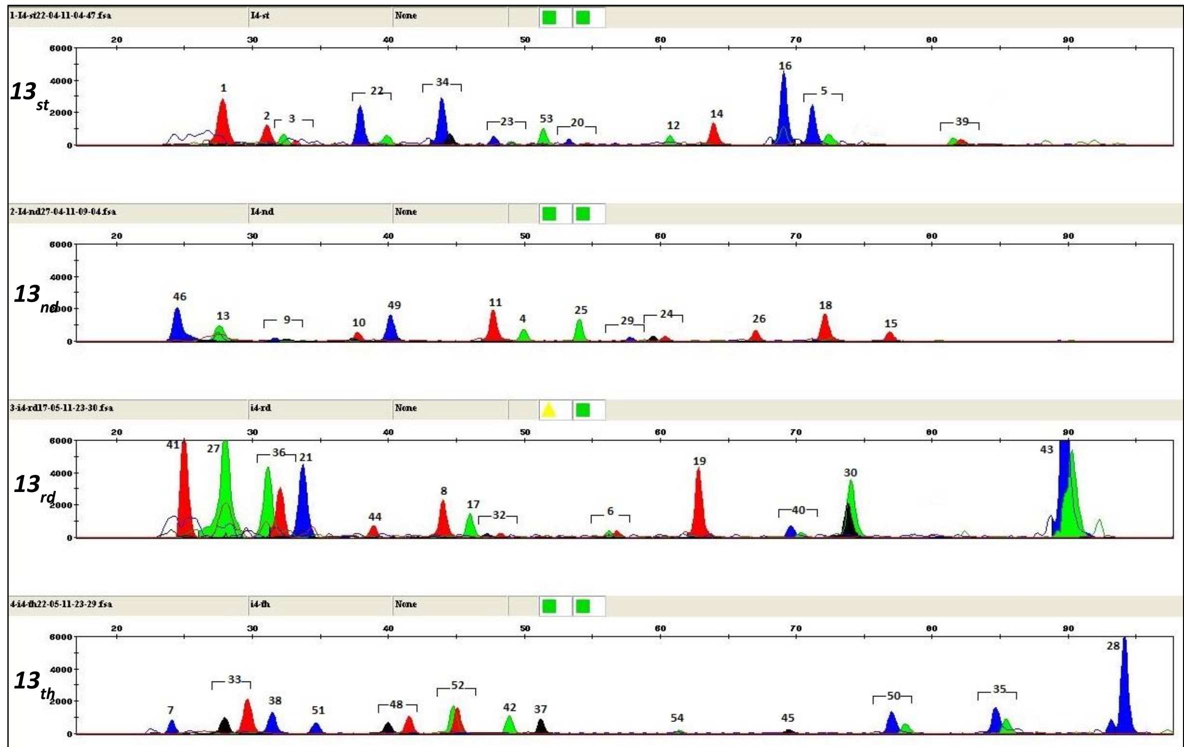


Figure 3-26: SNP profile from an individual with 1 µl (70 ng) of DNA template.

In Figure 3-27, arrows showed the marker locus where both alleles dropped-out. The samples with high DNA concentrations that gave poor SNP profiles, were diluted to 5 ng/µl before subjected to re-amplification and CE using AB 310 genetic analyzer.

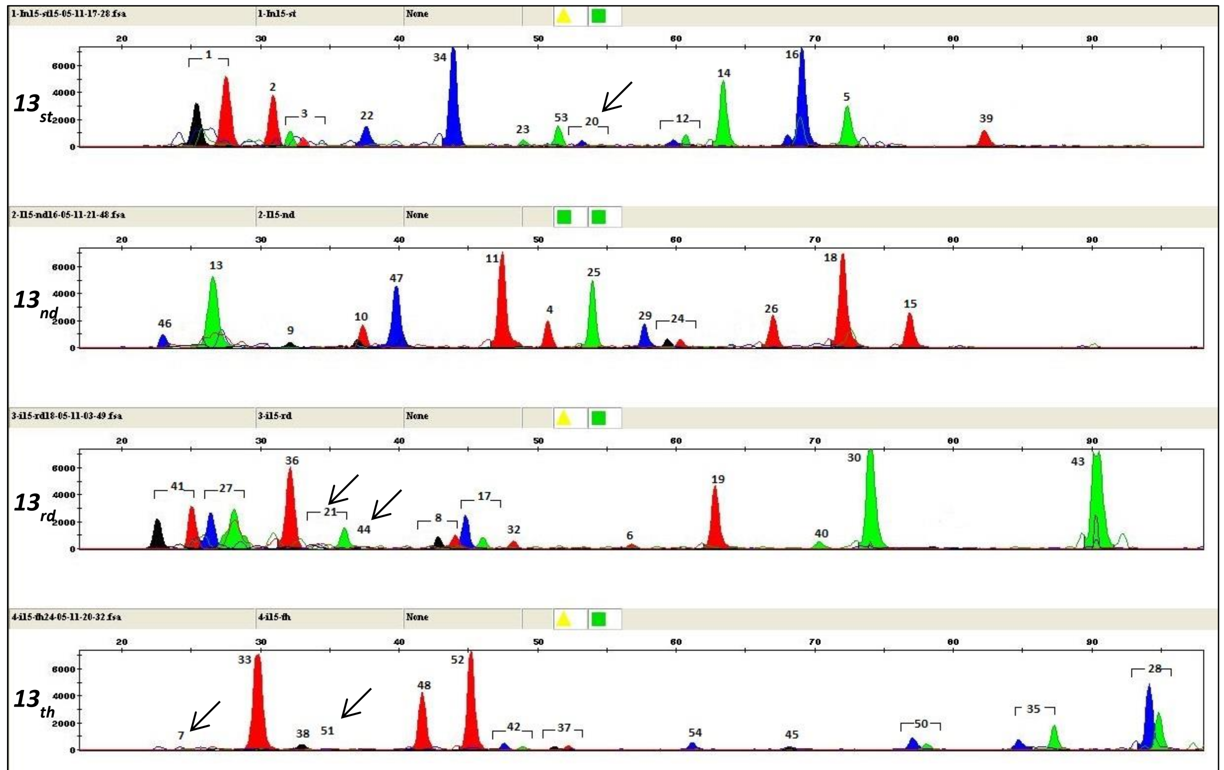


Figure 3-27: SNP profile of an individual with 1 µl (569 ng) of DNA template.

As for comparison to the TaqMan Genotyping Assays data, all tested samples that obtained from the SNP analysis showed identical allele genotypes as presented in Table 3-7.

Table 3-7: A summary of alleles genotyped in 90 Malaysian individuals using TaqMan® SNP Genotyping assays which similar to the SNaPshot assay data.

SNP Marker	rs1493232 (A/C)			rs1335873 (A/T)		
Ethnic	**Malay**	**Chinese**	**Indian**	**Malay**	**Chinese**	**Indian**
Type of Genotypes		6 AA	14 AA	16 AA	12 AA	19 AA
	21 AC	13 AC	11 AC	13 AT	13 AT	8 AT
	9 CC	11 CC	5 CC	1 TT	5 TT	3 TT

3.4. DISCUSSION

3.4.1. Selection of SNP markers and platform

To obtain a parallel power of discrimination as STR analysis in forensic casework more than 50 SNP markers are needed. However, currently the greatest problem in developing useful SNP arrays for forensic use is not related to statistical issues; rather, the problems are biochemical. Making a large balanced multiplex of 50 loci for less than 1 ng of genomic template is challenging (Gill 2001).

As stated by Sobrino, there are a large number of SNP genotyping methods available (Sobrino et al. 2005b), and usually the choice of the appropriate method depends on the number of SNPs and the number of individuals that need to be typed. Brion and colleagues have explained that the main reason favouring multiplexing PCR with the single base extension (SBE) reaction method using the SNaPshot multiplex kit for laboratories involved in forensic DNA analysis, was the possibility of genotyping SNPs without investment in new technologies, simply using a capillary electrophoresis (CE) instrument, which is available in most of these laboratories for multiplex STR typing by fragment analysis (Brión et al. 2005).

Therefore, in this study, 52 SNP markers identified from The SNP Consortium (TSC) and Sanchez et al. were selected to carry out SNP genotyping on Malaysian population and forensic casework samples. Sanchez and colleagues have established one PCR multiplex of 52 SNP markers with the two sets of multiplex system comprised of 23 SBE and 29 SBE reactions for human SNPs genotyping (Sanchez et al. 2006a).

3.4.2. Development of the multiplex assays

Multiplex PCR is a key technology for detecting infectious microorganisms, whole genome sequencing, forensic analysis, and for enabling flexible yet low-cost genotyping. However, the design of a multiplex PCR assay requires the consideration of multiple competing objectives (Rachlin et al. 2005b). As suggested by Butler, to develop an efficient multiplex PCR, careful planning, several tests in primer designing and balancing reactions are required (Butler 2011). In addition, Dixon et al. (2005) has stated that ppreparing new multi-mix batches was one of the main challenges as each batch will be slightly different and at the same time these new batches must be balanced. Many rounds of altering primer concentrations was also required to optimize inter or intra locus balance (Dixon et al. 2005). So, as for this study, several trials, such as developing 3-plexes, 5-plexes, 10-plexes and 13-plexes assays had been carried out along with SNaPshot single reaction of each marker individually, to observe and understand the primer-primer interactions and the capability of the SNP markers to be multiplexed.

The trial results showed that the combination and the concentrations of the PCR-SBE primers and the SBE product's size were very important in order to get clear and balanced alleles in a SNP profile. Nicodeme and Steyaert (1997) suggested electrophoresis distance constraints that require two amplicons to have some minimum length difference so that they can be distinguished after being processed through an electrophoresis gel (Nicodeme 1997).

The study from the original paper (Sanchez et al. 2006a) indicated that the sensitivity of ABI 310 PRISM® was poorer than the sensitivity of ABI 3100 PRISM®, and most of their studies were carried out using ABI 3100 or ABI 3100-Avant PRISM® Genetic Analyzers. Therefore, in this study, four reliable 13-plex assays (designated as *13_{st}, 13_{nd}, 13_{rd} and 13_{th}*) were developed from these 52 SNP markers to be analyzed with the ABI 310. This has improved the capabilities of the assays using ABI 310 genetic analyzer.

This final multiplex assays then were chosen to be used throughout the study. Each multiplex assay has 13 loci in the form of the lowest to the highest SBE amplicon size.

Fourteen SBE primer lengths were adjusted to balance the designated assays. The experimental results in this study have showed that the reduction of the primers (marker loci) used in one multiplex assay, as expected, improved the resolution of SNP profiles that were obtained. The optimal final combination and concentrations of the PCR-SBE primers were presented as in Table 3-6; together with the original established multiplex (Sanchez et al. 2006a).

Rachlin and colleagues have studied primer-primer interaction for screening individual primers as well as forward-reverse primer pairs, and for determining compatibility of two SNPs within a single multiplex PCR assay. They found that primer selection stringency is obviously a critical factor impacting the performance of any multiplex PCR design process. They have suggested that if fewer SNP markers more carefully chosen, it would be more likely to produce a working assay but could undermine one's ability to identify high multiplex designs (Rachlin et al. 2005a).

Another important factor that was taken into consideration was the optimal conditions of both, PCR and SBE reactions, such as the concentrations of PCR buffer, DNA Taq Polymerase, Mg^{2+} ions, deoxynucleotides or dideoxynucleotides and the PCR/SBE cycling conditions. Several adjustments of PCR/SBE components' concentrations were tested using the same thermal conditions as described by Sanchez et al. (2006a). The final optimal compositions of the PCR/SBE reactions which produce balanced and good SNP profiles in all 13-plex assays were described in *Chapter 2, sections 2.7.2.3. to 2.7.2.8.1.*

Also, it is very important to ensure that the correct targeted SNP allele(s) were obtained and the appropriate dyes exhibited by 13 SNPs homozygotes or heterozygotes alleles in each multiplex assays. Results from the experiment have showed that the targeted SNP allele(s) of each SNP marker were correctly labelled with the correct dyes, and they exhibited stably for all four 13-plex assays. This shows that the reduction of alleles loci in one multiplex assay, has decreased the dyes and primer-primer preferential amplification and makes the interpretation of the profile became easier. Borsting et. al. (2009) stated that the strengths of the fluorophore emissions from the four dyes used in the SNaPshot

kit are unbalanced and that the peak height of one allele may be up to 6 times higher than the peak height of the other allele in the same locus. This scenario increases the risk of false interpretation of the results by the analyst, especially when many SNPs need to be analyzed in the same electropherogram (Borsting et al. 2009).

The analysis and interpretation of the SNP profiles have generally relied on the observation of the number and dye colour(s) of the peaks at each locus and the relative allele peak heights. Therefore, the multiplex assays developed in this study proposed a simpler and easier alternative to prepare and genotype the samples. It is also suitable for the laboratories that only have ABI 310 genetic analyzer to carry out genotyping analysis.

Furthermore, clear resolution and full profile of 52 SNPs were observed from all series of control DNA dilutions. As described by Borsting et al.(2009) that when some discrepancies were observed between the laboratories, it was often caused by simple misinterpretations of the electropherograms such as the SBE primers being labelled with two different dyes, for the heterozygous individuals, and are detected at different wavelengths and appear in different dye windows of the electropherogram.

3.4.3. The stability of the multiplex assays

Along with the control DNA, TaqMan DNA and M7-7, 90 test population samples were used to study the reproducibility and stability of these four 13-plex assays, where 1 µl of each sample, where template DNA ranged from 0.04 ng up to 569 ng were used. The results obtained show the sensitivity and reproducibility of the 13-plex assays developed, which they were able to amplify DNA as low as 0.03 ng and higher than 70 ng. As for forensic casework, these 13-plex assays could be useful tools in detecting low template DNA that less than 0.2 ng. In addition, the concordance TaqMan genotyping data confirmed that the targeted alleles were correctly genotyped using the 13-plex assays.

Sanchez et al. (2006b) has reported that the multiplex PCR (with 23 and 29 SBE reactions) has allowed full complete allele-called from DNA input as low as 0.5 ng. However, the optimal amount of DNA in the PCR was 1–10 ng for successful amplification, and acceptable results were obtained with up to 70 ng DNA. With DNA input lower than 0.2 ng, the alleles started to drop-out and gave unbalanced peak heights.

CHAPTER 4

VERIFICATION STUDIES: ANALYSIS OF FOUR 13-PLEX SNAPSHOT ASSAYS USING ABI PRISM® 3500 GENETIC ANALYZER

ANALYSIS OF 13-PLEX ASSAYS (13_{st}, 13_{nd}, 13_{rd} and 13_{th})

USING 3500 GENETIC ANALYZER

4.1. INTRODUCTION

Over the last years there has been a vast number of developments and applications of capillary electrophoresis (CE) and microchip-CE. This growth has influenced the recent developments in instrumentation associated with CE. CE offers high separation efficiency, low cost, fast analysis, and minimal waste generation (Felhofer et al. 2010). Today, CE and microchip-CE devices have become well accepted, multi-dimensional analytical platforms in the biomedical, pharmaceutical, environmental, and forensic sciences.

CE is the primary methodology used for separating and detecting short tandem repeats (STR) alleles in forensic DNA laboratories worldwide. The first CE introduced was the single-capillary ABI PRISM® 310. But currently, a number of capillary array electrophoresis (CAE) systems are commercially available (Butler et al. 2004). These CAE systems offer from 8 to 384 capillaries run in parallel for the high throughput analysis as shown in Table 4-1.

The latest generation of CE instruments are 3500 and 3500xl Genetic Analyzers, which became commercially available in 2010, supplied by Applied Biosystems (Life Technologies, Foster City, USA). These instruments are available in 8 capillary (3500) and 24 capillary (3500xl), and were designed with simplified setup reaction, operation and maintenance. These series are also incorporated with the RFID (Radio Frequency Identification) labels on all capillary arrays, polymer pouches, buffer containers and conditioning pouches. The RFID labels are for tracking and reporting the consumable usage (lot, part numbers and expiration dates). The consumables data can be retrieved or tracked from Data Collection Software even after the consumables have been removed from the instrument. These 3500 series designed for use with the 96 or 384-well microtiter plates' format and also for 96-well Fast and 8-tube standard or fast

strips (Applied Biosystems 2009b). Other additional features includes an improved mechanical pump for polymer filling of the capillaries with less waste, new laser technology, reduced power requirements and 6-dye detection capabilities (Butler 2011).

Table 4-1: The number of arrays in commercial CAE systems [Adapted from (Butler et al. 2004, Butler 2011)].

Type of Arrays	No. of capillaries
Applied Biosystems (Foster City, CA, USA)	
ABI 3100 Avant	4
ABI 3100	16
ABI 3130 series	4 & 16
ABI 3700	96
ABI 3730 series	48 & 96
ABI 3500 series	8 & 24
Amersham Biosciences (Piscataway, NJ, USA)	
MegaBACE 500	48
MegaBACE 1000	96
MegaBACE 4000	384
SpectruMedix Corporation (State College, PA, USA)	
SCE 2410	24
SCE 9610 96	96
SCE 19210	192
Beckman Coulter (Fullerton, CA, USA)	
CEQ 8800	8

4.2. RATIONALE OF THE STUDY

As demonstrated in the previous chapter, four sets of 13-plex PCR/SBE assays (*13_{st}, 13_{nd}, 13_{rd} and 13_{th}*) have been developed and internal validation study includes the sensitivity and reproducibility assessment had been carried out using the ABI PRISM® 310 genetic analyzer. In addition, 90 population samples have been evaluated.

4.2.1. Concordance analysis using the 3500 Genetic Analyzer

The main aim of this study was to assess the sensitivity and reproducibility of the *13_{st}, 13_{nd}, 13_{rd} and 13_{th}* assays using the 3500 Genetic Analyzer before applied these assays to a larger number of population samples and forensic casework samples.

To begin with, all 13-plex assays generated from each series of Control DNA from the previous study (*Chapter 3, Part Two*), were re-electrophoresed using a 3500 Genetic Analyzer with polymer POP-7™ as recommended by the manufacturer's protocol (Applied Biosystems 2009b).

4.2.2. Assessment of the 13-plex assays using different polymers and parameters

Following the primary results obtained, a few experiments were carried out to test the three commercially available polymers for SNP analysis on the 3500 genetic analyzer. Applied Biosystems supplied two primary polymers, that for use with the ABI PRISM 310 and other CAE systems: POP-4™ and POP-6™, where "POP" is stands for Performance Optimized Polymer. POP-4™ is commonly used for DNA fragment analysis while POP-6™ is capable of higher resolution, which is suitable for DNA sequencing. Recently, POP-7™ has been also introduced for both applications (Butler 2011).

POP™ polymers dynamically coat the capillary wall to control electro-osmotic flow as illustrated in Figure 4-1. This figure illustrates a cross-section through a CE capillary. Electrophoretic flow is the movement of DNA fragments from the negative electrode

(cathode) towards the positive electrode (anode). There are strands of entangled polymer that sieve these DNA fragments progress based on their size, meaning that the smaller fragments will get through the obstructions easier than the longer fragments.

There is also another opposite force that also produced inside the capillary, which known as an electro-osmotic flow or EOF. The capillaries are made of glass or fused silica (silicon dioxide, SiO_2) which has a negatively charged hydroxyl group (silanol, SiOH) along the inner walls. Under the electric field strength, the positive ions from the electrophoresis buffer will move towards the negative electrode producing a plug like flow having a flat velocity distribution across the capillary diameter. Velocity of the electro-osmotic flow depends on the composition, pH and ionic strength of the electrolyte solution (Butler 2011, Corradini, Sprecacenere 2003).

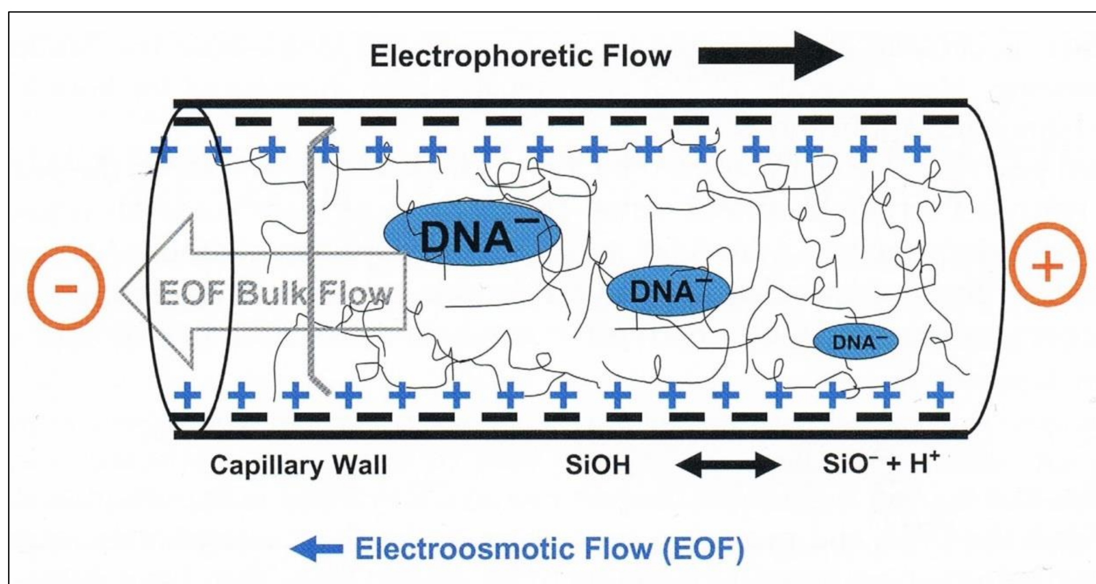


Figure 4-1: Illustration of internal capillary environment and forces with moving DNA molecules through a sieving polymer buffer. [Taken from (Butler 2011)].

Madabhushi and colleagues (1997, 1998) studied Poly (N,N-dimethylacrylamide)/ pDMA as a DNA-sequencing matrix, focusing on this self-coating feature. However, ABI holds a basic patent on any water soluble polymers with self-coating features. These polymer comprising 4% and 6% concentrations of dimethyl polyacrylamide and high concentration urea (Wenz 1994, Rosenblum et al. 1997, Madabhushi et al. 1997, Wenz et al. 1998, Lazaruk et al. 1998, Madabhushi 1998, Maruyama 2005, Butler 2011).

Since ABI declared that POP, the polymer matrices for its sequencing instruments, has self-coating features, it has been widely believed that POP consists of pDMA. ABI has never released any part of POP's ingredients, except from the related published articles that stated the 3700™ instrument uses a separation matrix that is based on relatively low molar mass, linear poly-dimethylacrylamide, which is commercially known as POP 7 (Maruyama 2005). Apart from the above information, the only characteristic of the polymers known is as listed Table 4-2.

Table 4-2: Polymer characteristics [Adapted from (Applied Biosystems 2012).

Polymer type	Characteristics
POP-4™	*Less viscous, fast runs*
POP-6™	*More viscous, slow runs*
POP-7™	*Less viscous, fast runs*

In order to carry out the polymer evaluation for genotyping the 13-plex assays, approximately 20 control and population samples were re-analyzed using a 3500 Genetic Analyzer but with different parameters as shown in Table 4-3.

Table 4-3: Run parameters for the SNP analysis on the 3500 Genetic Analyzer.

3500 Genetic Analyzer Run Parameters			
Parameters	**Polymer types**		
	POP-4™	**POP-6™**	**POP-7™**
Capillary array	*36 cm*	*50 cm*	*50 cm*
Application type	*HID*	*Fragment*	*Fragment*
Dye set	*E5*	*E5*	*E5*
Run module	*HID36_POP4*	*FragmentAnalysis50_POP6*	*FragmentAnalysis50_POP7*
Run temperature	*60°C*	*60°C*	*60°C*
Injection voltage	*1.2 kV*	*1.6 kV*	*1.6 kV*
Injection time	*10 s*	*10 s*	*10 s*
Run time	*2800 s*	*2800 s*	*2800 s*

4.2.3. Analysis of tested samples with optimal parameters and conditions

Based on the preliminary results obtained, which displayed over-amplification, the volume for each SBE multiplex assay was reduced to 1.7 µl of SNaPshot mix, 1 µl of purified PCR product and 1 µl of mixed SBE primers, and made to a final volume of 5 µl with nuclease-free water.

As for final optimization, all control DNA samples (single and multiplex reactions), with template amounts ranging from 0.01 ng to 2.5 ng (in triplicates) and tested population samples were analyzed with polymer POP-6™ on a 3500 Genetic Analyzer as described in *Chapter 2, section 2.7.2.8.2.*

4.3. RESULTS

4.3.1. Separation and detection of 13-plex assays on the 3500

Initially, the analysis of the same SBE products (from the previous internal validation study-*Chapter 3, Part two*) was carried out using POP-7™ with 50 cm capillary array on the 3500, following the recommended manufacturer's protocol (Applied Biosystems 2009b). However, the resolution and mobility aptitude of the alleles was very poor, especially for the SBE products below 60 nt. As shown in Figure 4-2(A), *13_{th}* assay profiles of Control DNA samples that with (0.03, 0.125, 0.25 and 0.5) ng demonstrates poor resolution. The first eight alleles in those profiles are close to each other.

A similar observation was noted in few other tested samples that had been re-analyzed on the 3500 genetic analyzer. As shown in Figure 4-1(B), a worse scenario was observed when compared to Figure 4-2(A), because these samples have more heterozygous SNP alleles. The *13_{rd}-plex* assay comprises alleles' ranging in size from 16 nt to 88 nt but on polymer POP-7™, the alleles' size have been shifted to approximately from 32 nt to 95 nt.

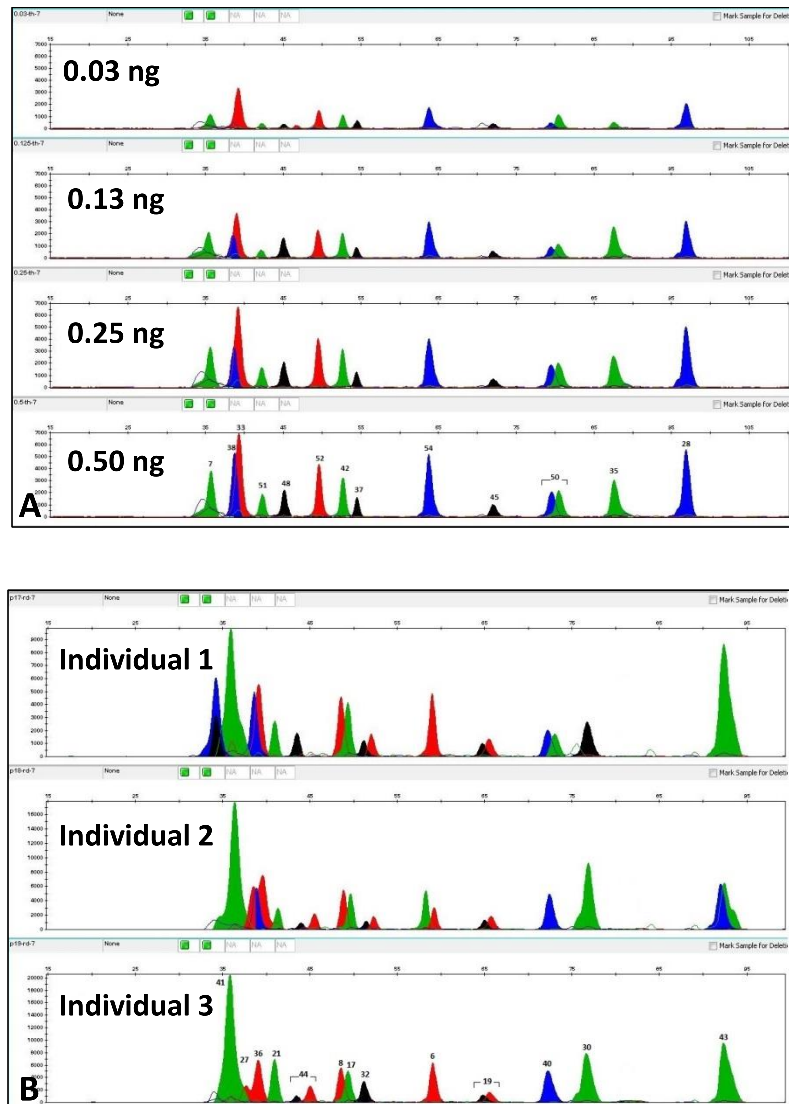


Figure 4-2: Electropherograms displaying the separation of: (A) SNaPshot profiles of 13_{th} multiplex assay of Control DNA, and (B) SNaPshot profiles of 13_{rd} multiplex assay of 3 different individual samples using POP-7™ on 3500 Genetic Analyzer.

4.3.2. Effect of different polymers and parameters on the 13-plex assays analyzed using the 3500 Genetic Analyzer

Re-analysis of the same SBE products but with different parameters and polymers, POP-4™ and POP-6™ was carried out as described in Table 4-3.

The results obtained were compared as shown in Figure 4-3, where the 13_{th}-plex assay profiles of the same SBE products (0.5 ng Control DNA) has been analyzed with 36 cm capillary and POP-4™ on ABI 310 (A), 36 cm capillary and POP-4™ on ABI 3500 (B), 50 cm capillary and POP-6™ on ABI 3500 (C) and 50 cm capillary and POP-7™ on ABI 3500 (D). It was clearly noted that this sample has more homozygous alleles that differ in size by at least more than 2 nucleotides (nt) from each other. But, each of the profiles (Figure 4-3 B, C and D) showed different allelic migration patterns, especially for SNP alleles below 60 nt. However, the peak heights exhibited were quite similar (5000 to 7000 RFU) for all profiles.

The boxes show the most affected area of the allele migrations. But, the separation of the SNP alleles obtained from sample runs on the 310 (A) and 3500 (B) analyzers have no significant differences. Sample on the POP-7™ has very poor resolution. However, the sample analyzed with POP-6™ (C) on the 3500 gave the best allele resolution.

The close-up of these selected areas is shown as in Figure 4-4 (A, B and C). The arrows show the allele migration from the SNP code marker’s locus that was created based on the allele’s size of Control DNA (triplicates) generated using POP-6™ polymer. The difference in allele size obtained for SNP markers 7, 33, 38 and 48 of these profiles were calculated from the actual allele size for each particular marker.

As can be seen from Table 4-4, there are huge differences between allele migrations in POP-7™ compared to POP-4™ and POP-6™ polymers. This data (Control DNA) was obtained from one same sample that has been electrophoresed using all polymers.

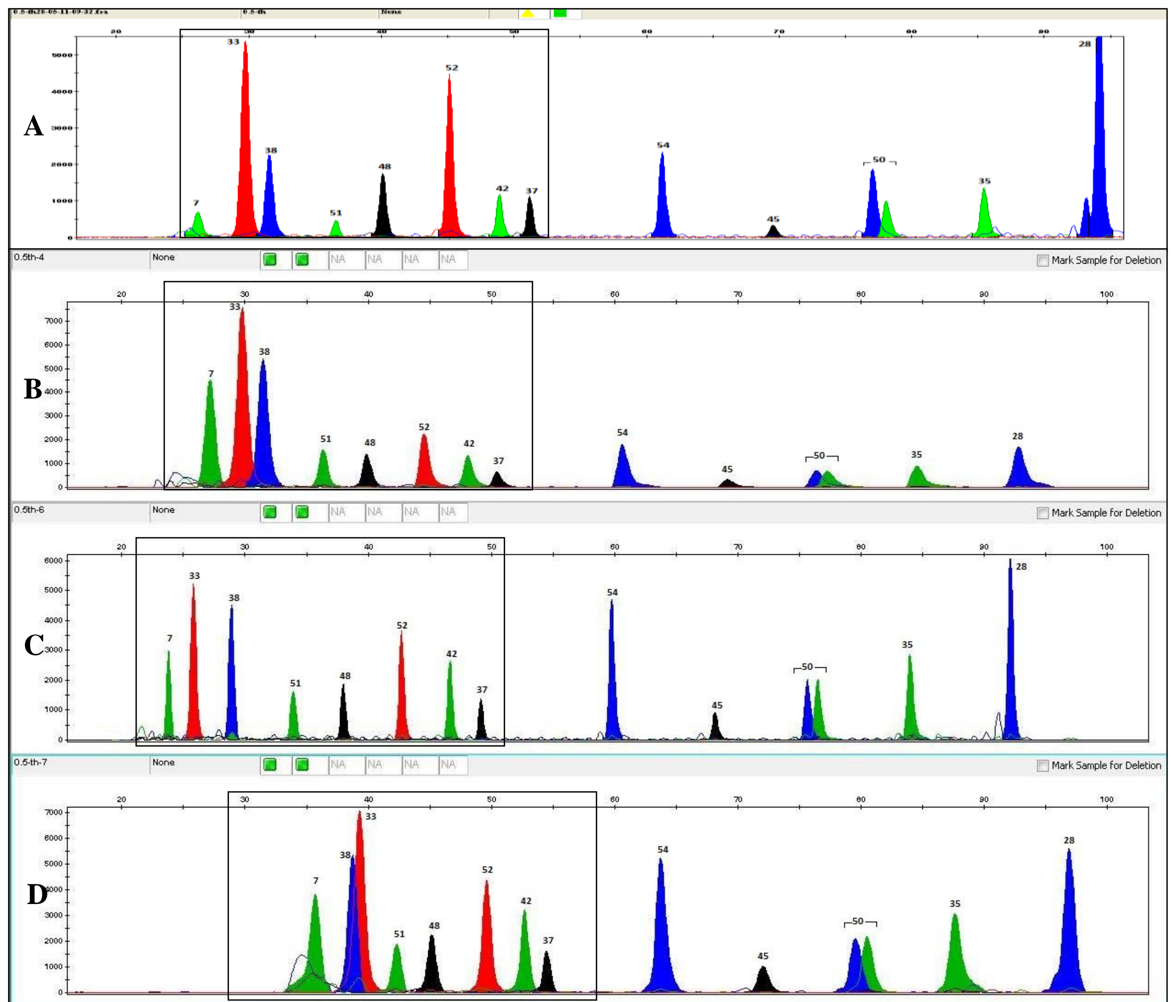


Figure 4-3: Comparison of SNaPshot profiles of 13_{th} multiplex assay using polymers; A) POP-4™ on ABI 310 PRISM® Genetic Analyzer, B) POP-4™, C) POP-6™ and D) POP-7™ on the 3500 Genetic Analyzer. Boxes show the most affected SNPs alleles.

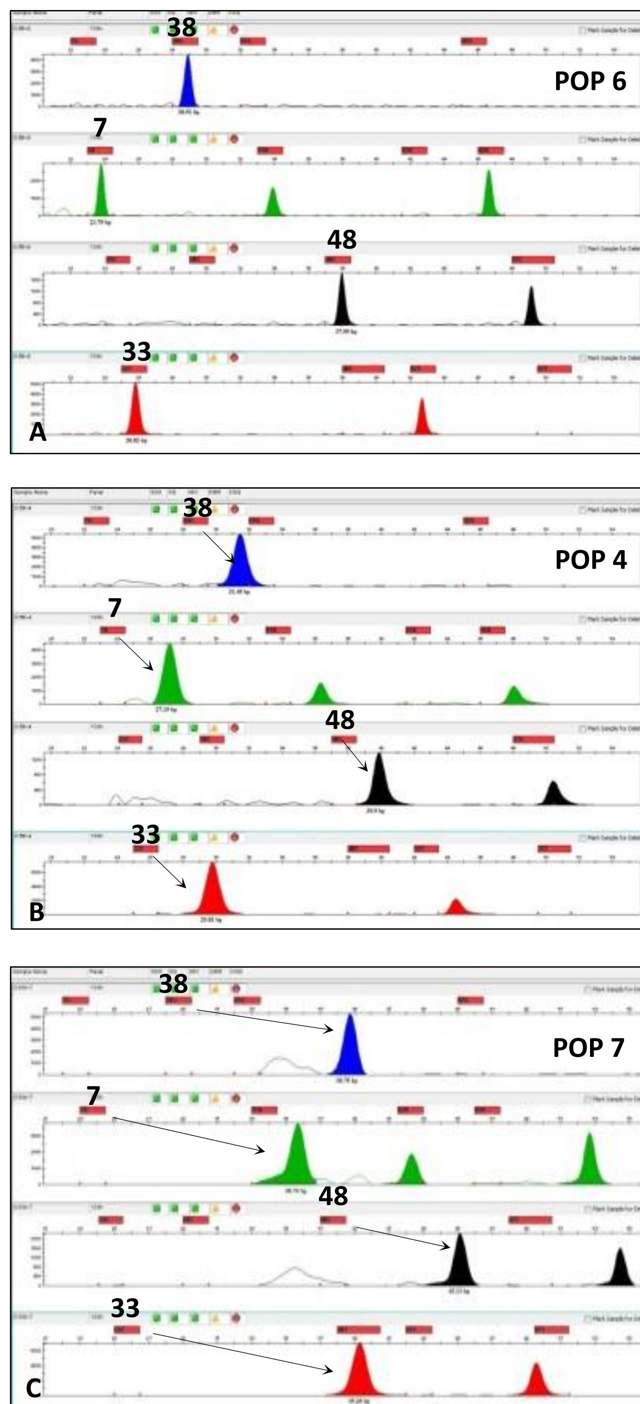


Figure 4-4: Allele migration of Control DNA in different polymer types; A) POP-4™, B) POP-6™ and C) POP-7™ on 3500 Genetic Analyzer. Code markers' positions were highlighted in red. The arrows show the first allele's movement in each dye.

Table 4-4: Data of allele's migration (bp) on different type of polymers.

Allele's migration measurement(nt)				
Markers	**Actual Size (nt)**	**POP-6™**	**POP-4™**	**POP-7™**
38	27.0	28.91	31.45	38.75
Size deviation		***1.91***	***4.45***	***11.75***
7	18.0	23.79	27.19	35.74
Size deviation		***5.79***	***9.19***	***17.74***
48	36.0	37.99	39.90	45.13
Size deviation		***1.99***	***3.9***	***9.13***
33	22.0	25.82	29.81	39.26
Size deviation		***3.82***	***7.81***	***17.26***

The polymer performance evaluation has also been carried out on the Positive Control provided along with the ABI PRISM® Multiplex Kit and other tested population samples. The results are shown as in Figure 4-5. As observed, the green peaks and a blue peak of the Positive Control have merged into one huge peak on the POP-7™ polymer. Whilst the separation on POP-4™ and POP-6™ can still differentiate all the alleles although there is a noticeable mobility differences between these two polymers.

The alleles were not separated well even for the population sample analyzed on the POP-7™. As can be seen from Figure 4-5 B, the first five alleles have merged as a complicated complex and it was hard to identify them correctly. This 13_{rd}-plex assay has actual allele size ranging from 16 to 88 nt, but for these five alleles alone, the size ranging from 16 to 38 nt. However, from the results obtained on POP-7™, the allele's size for markers 41, 27, 36, 21 and 44 was cramped from 34.28 to 43.47 nt. On POP-6™, these five alleles have better separation with allele sizes ranging from 19.47 to 35.36 nt.

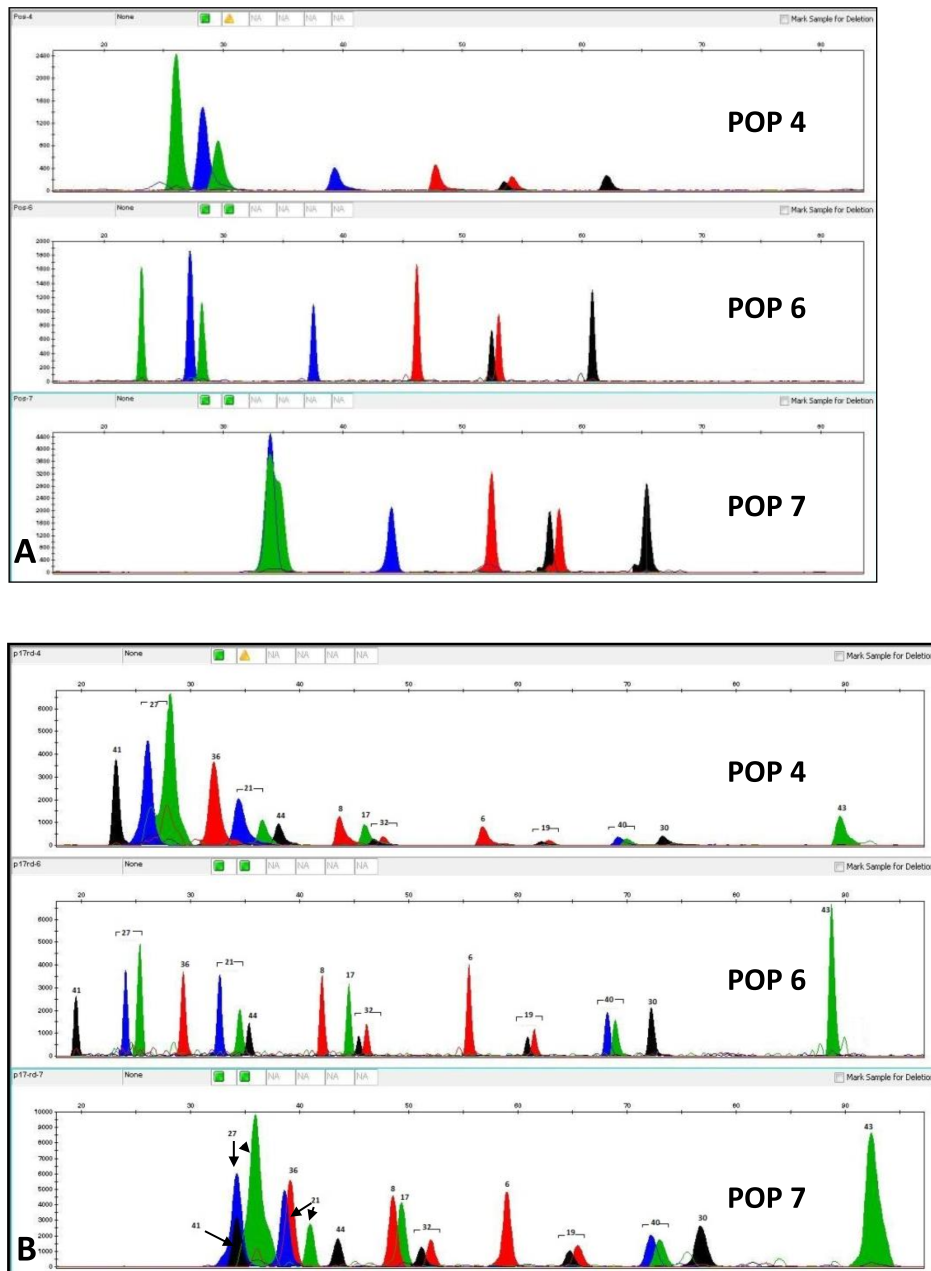


Figure 4-5: (A) SNP profiles of Positive Control of ABI PRISM® SNaPshot Multiplex kit and (B) 13_{rd}-plex profiles of an individual sample electrophoresed in different polymer types; POP-4™, POP-6™ and POP-7™ on 3500 Genetic Analyzer. Arrows show poor separation of the allele's occurrence.

4.3.3. Sensitivity and reproducibility assessment using polymer POP-6™ and optimal parameters on the 3500 Genetic Analyzer

4.3.3.1. Detection level of sample loaded on the 3500

To observe SNP allele's detection level on the 3500 genetic analyzer, 0.4 µl and 1 µl of each type of 13-plex assays were loaded into 96 microtitre-plates, together with the internal size standard and formamide. Figure 4-6 demonstrates that no significant difference was observed between the profiles' peak height and profiles' quality for Figure 4-6 (A) and (B), but as for (C) and (D), a slight difference in the peak height was observed. This scenario was witnessed only in longer assays: *13_{rd}*-plex and *13_{th}*-plex profiles, which the alleles' size approximately around 16 to 92 nt, compared to the other two shorter 13-plex assays, the alleles' size were from 16 to 80 nt.

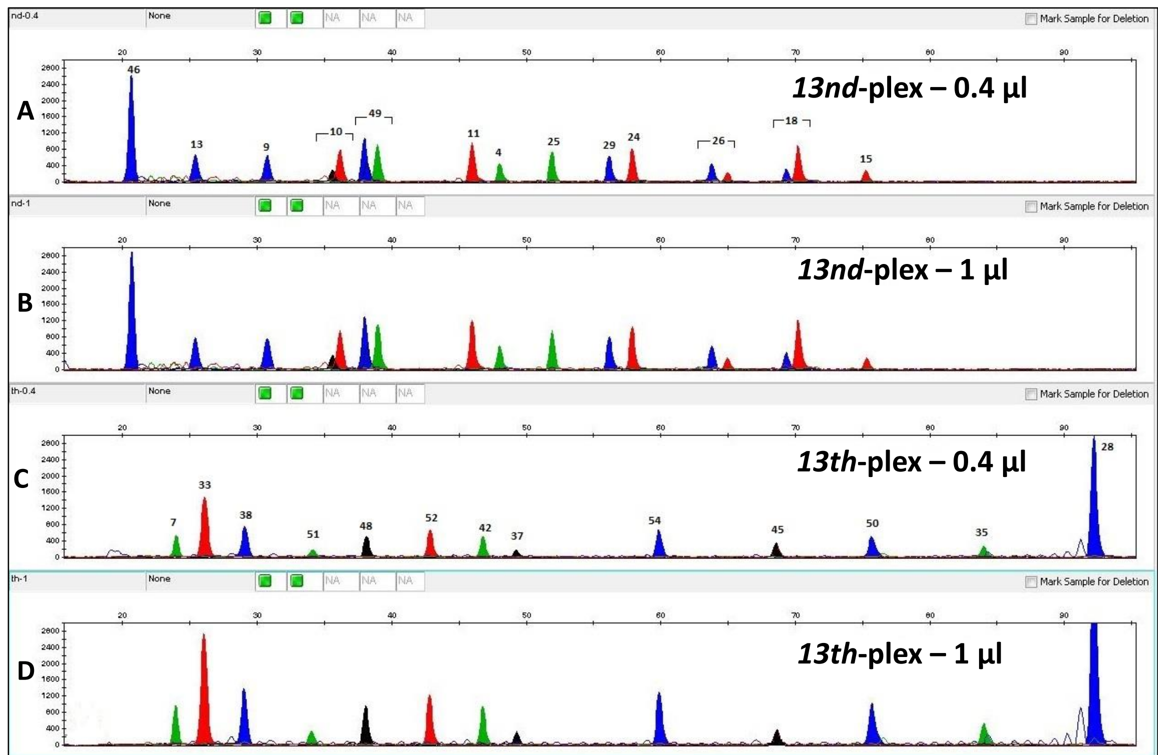


Figure 4-6: A) and B) are the profiles of *13_{nd}*–plex assay, but with 0.4 µl and 1 µl of samples loaded, respectively. C) and D) are the profiles of *13_{th}*-plex assay, with 0.4 µl and 1 µl of samples loaded, respectively using polymer POP-6™ on the 3500 Genetic Analyzer.

4.3.3.2. Peak height observation

Results obtained from the concordance study showed that the peak height in relative fluorescence units (RFU) of the same product was 3 times higher when it was electrophoresed using the 3500 (Figure 4-7(B)) than the ABI 310 (Figure 4-7(A)), which were 14,000 RFU and 4200 RFU, respectively. Because of these observations, the SBE reagents consumed for multiplex assays have been reduced as mentioned in *section 4.2*. Figure 4-7(C) shows the same PCR product generated using the new optimal SBE conditions (reduced reaction mix) that has exhibited average peak height (5000 RFU) on the 3500. This determines that the sensitivity and detection limit of 3500 is higher than ABI 310.

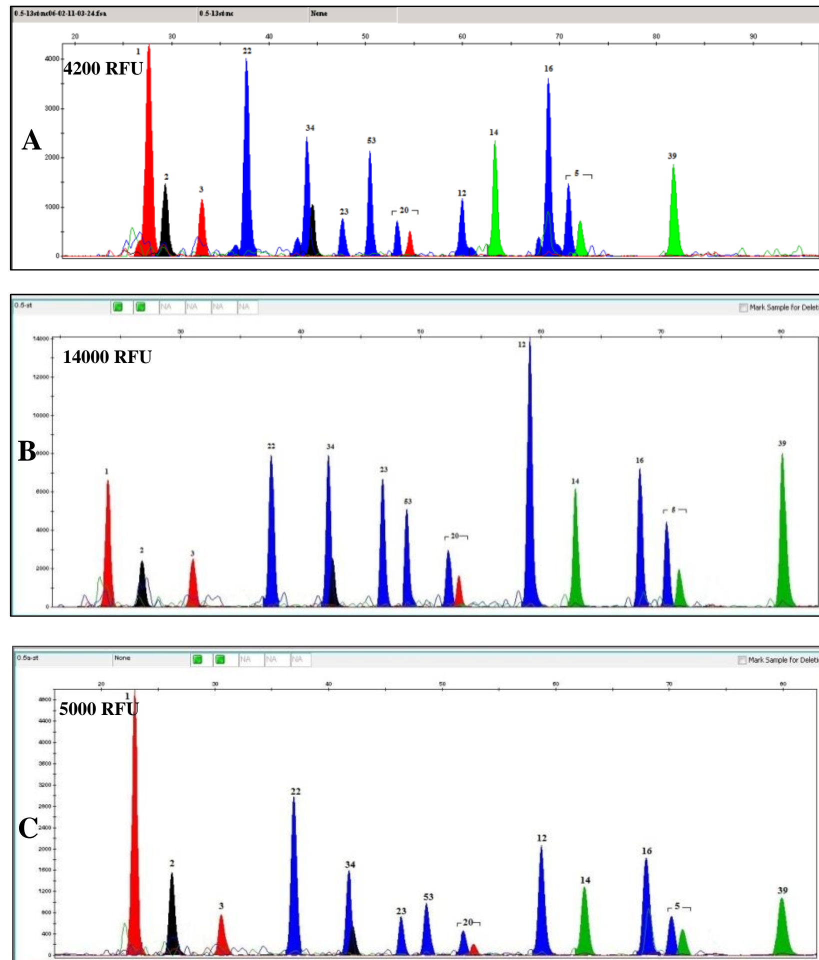


Figure 4-7: Peak height (RFU) comparison of three SNP electropherograms of the same Control DNA sample (0.5 ng/µl) using; (i) an optimal conditions and parameters with ABI 310 PRISM® Genetic Analyzer, (ii) Same product as (A) but electrophoresed and analyzed with 3500 Genetic Analyzer, and iii) new optimal conditions with 3500 Genetic Analyzer.

Above results have indicated that the new optimal criteria of the capillary electrophoresis using 3500 Genetic Analyzer were the PCR and SBE reactions and conditions as mentioned in section 2.3 and 2.7. Whilst the suitable polymer was POP-6™, with 50 cm capillary array, 10s (injection time), 1.6 kV (injection voltage), 2800s (run time), E5 dye set and with run module: fragmentAnalysis50_POP6.

4.3.3.3. Stability of the assays on the 3500 Genetic Analyzer

With these new optimal conditions and parameters, several samples including the controls (TaqMan Control DNA and 9947A) have been re-analyzed using the 3500. All Control DNA samples which ranged from 0.01 to 2.5 ng were analyzed in triplicate for comparison purpose. Figure 4-8 shows triplicates of two different DNA concentrations (2.5 and 0.1 ng) of two different assays (13_{th} and 13_{nd} assays). Results determined the stability of the assays.

4.3.3.4. Sensitivity level of DNA input

As for the sensitivity of the assays on the 3500, a full SNP profile (52 markers) was obtained from all Control DNA and samples with input higher than 30 pg. In some samples with 30 pg template DNA one or two allele(s) dropped out. In addition, peak imbalances and additional peaks were observed for DNA input lower than 30 pg. Overall, from the results obtained, these 13-plex assays found to work best and consistently with template DNA of 60 pg and higher.

Figure 4-9 shows the comparison of 13_{nd}-plex profiles from different DNA concentrations: (0.01, 0.02, 0.03, 0.06, 0.125, 0.25 and 0.5) ng. Full 13 alleles are clearly observed from DNA amount of 30 pg to 0.5 ng. Well-balanced profiles were witnessed from samples with DNA input of 60 pg to 0.5 ng.

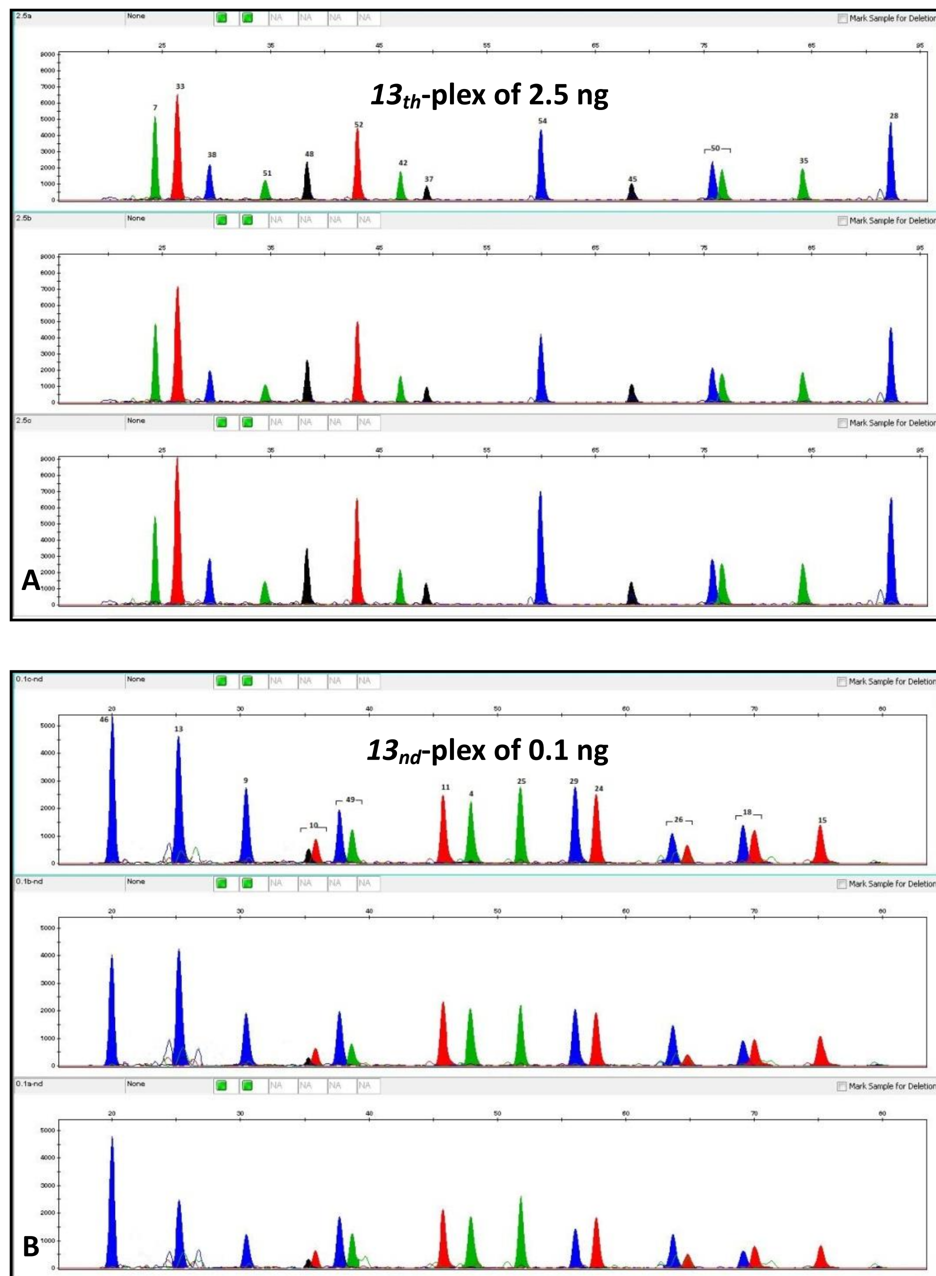


Figure 4-8: Electropherograms of; A) triplicates of *13th*–plex assay with 2.5 ng of Control DNA and B) triplicates of *13nd*-plex assay with 0.1 ng of Control DNA, using polymer POP-6™ and 3500 Genetic Analyzer, each at constant 9000 RFU and 6000 RFU, respectively.

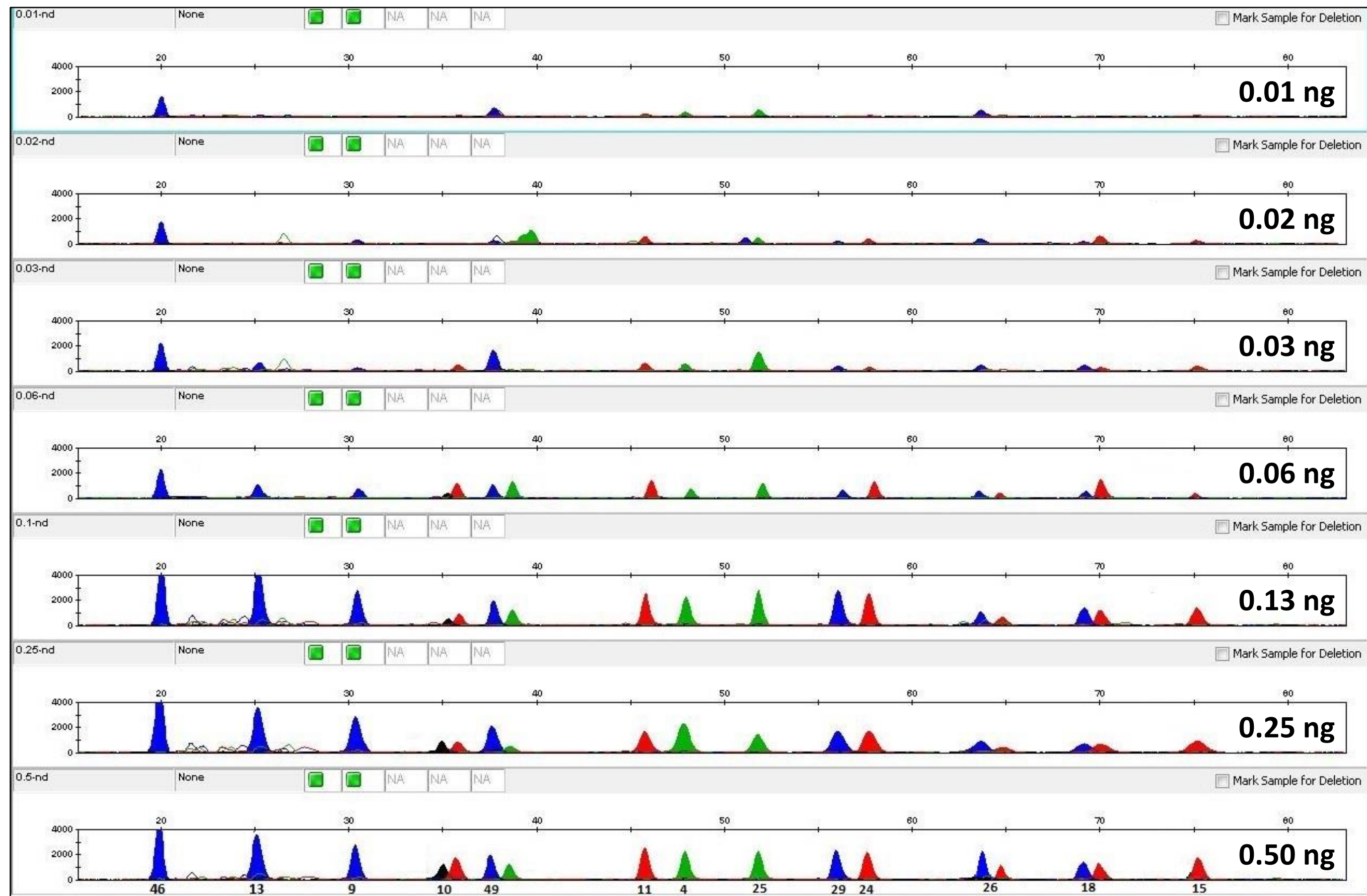


Figure 4-9: *13_{nd}* -plex profiles of a series of DNA template input: all are shown with a scale of 4000 RFU.

The potential of these assays in detecting SNP alleles using the 3500 genetic analyzer under controlled laboratory conditions has been tested with focusing on the DNA sensitivity level that is from 10 pg to 60 pg of DNA input. As shown in Figure 4-10, at very low levels of DNA template stochastic variations such as allele drop-in, allele drop-out and peak imbalances are seen. The triplicates of these DNA template amounts were consistent.

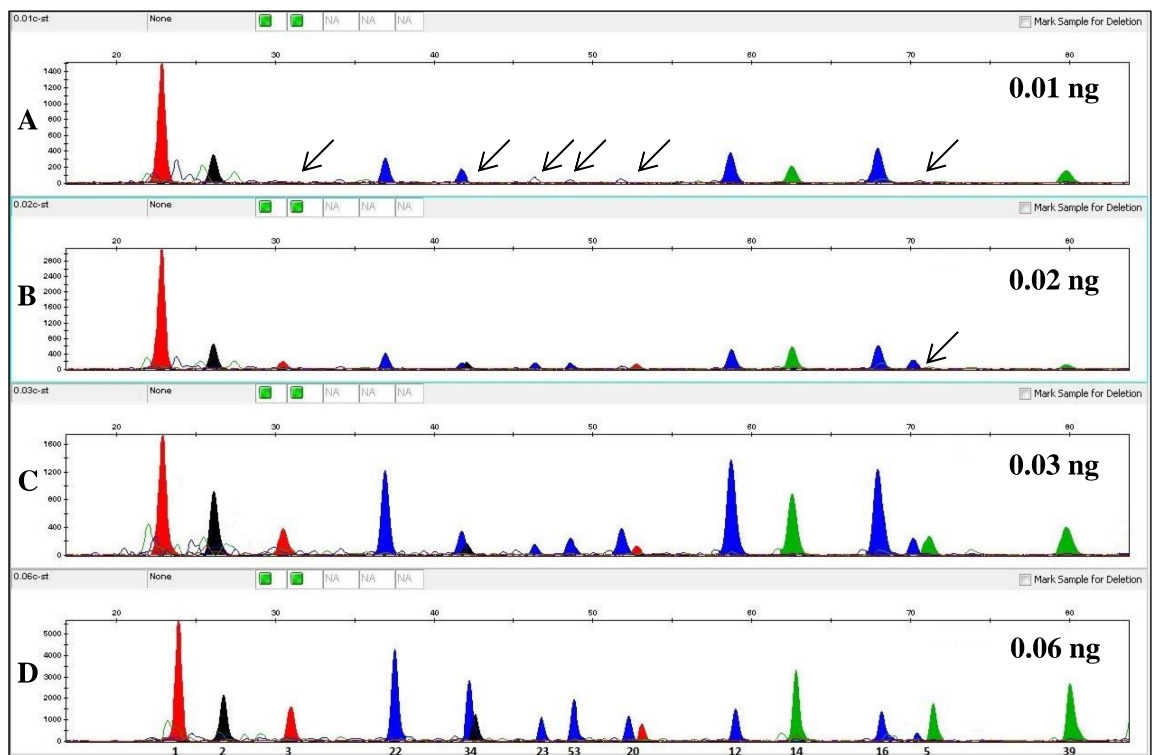


Figure 4-10: SNP electropherograms of 13_{st}–plex assay generated from DNA serial dilutions: A) 0.01 ng, B) 0.02 ng, C) 0.03 ng and D) 0.06 ng. The arrows show the targeted allele(s) that dropped out.

Lower template levels were also seen to improve results, overcoming some bias problems of the preferential amplifications and electrokinetic injection that were seen with higher template levels. As shown in Figure 4-11, the same sample with two different DNA amounts: 3.3 ng and 0.3 ng, were analyzed using the same optimal conditions have demonstrated different quality of SNP profiles. The sample with lower input DNA has better resolution with a cleaner baseline.

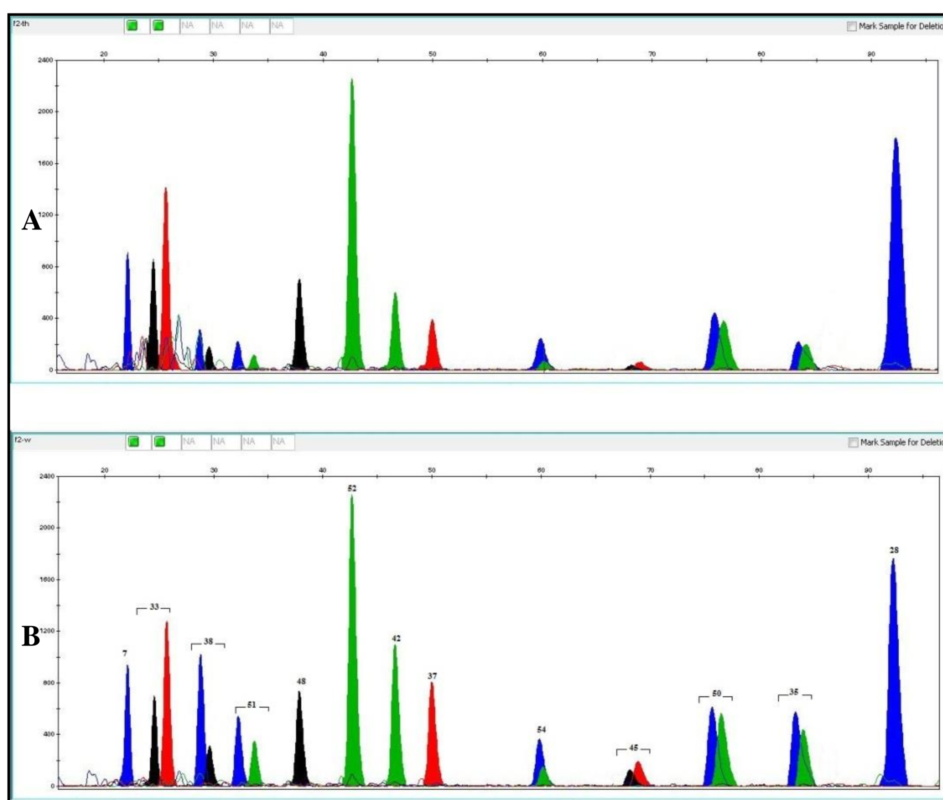


Figure 4-11: SNP profiles generated from the same sample but with different DNA concentrations; A) 3.3 ng/µl, and B) 0.3 ng/µl, electrophoresed with polymer POP-6™ using ABI 3500 PRISM® Genetic Analyzer.

All four 13-plex assays were arranged as 4 panels per individual (52 autosomal SNP markers) for better visualization and interpretation of the SNPs profile. Figure 4-12 demonstrates a full SNP profile (13_{st}, 13_{nd}, 13_{rd} and 13_{th}) obtained from 1 ng of 9947A on the POP-6™ and 3500 Genetic Analyzer.

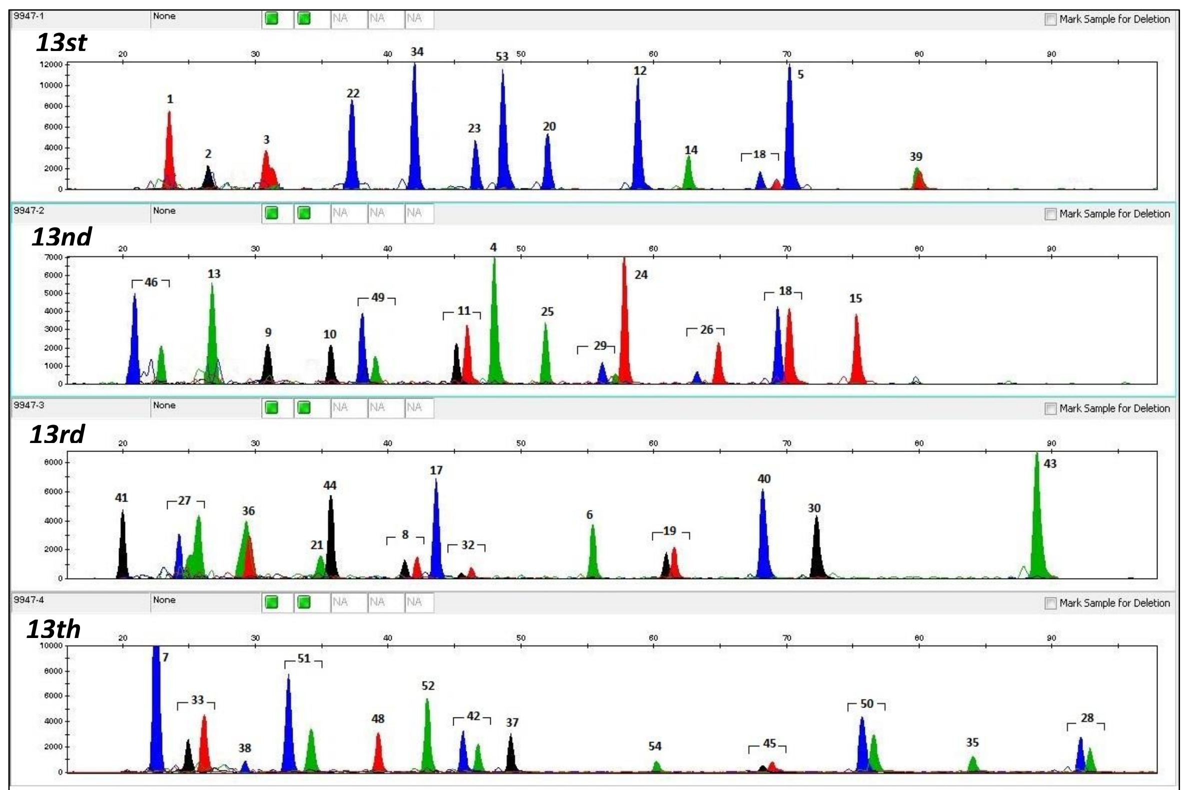


Figure 4-12: SNPs electropherogram of 52 markers generated from 1 ng 9947A.

Another example of full SNP markers in one electropherogram that generated from tested population sample.

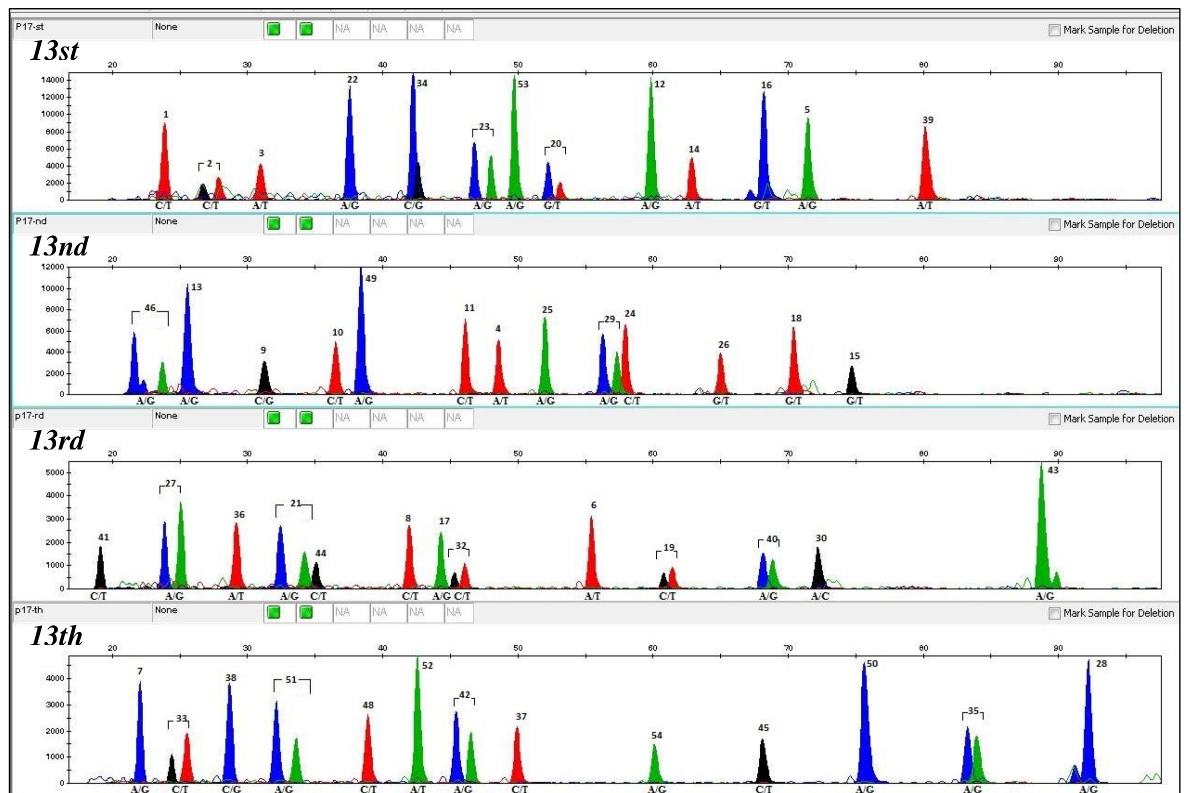


Figure 4-13: SNPs electropherogram of 52 markers generated from 1 ng tested individual sample.

4.4. DISCUSSION

In the previous chapter, four reliable 13-plex assays (designated as 13_{st}, 13_{nd}, 13_{rd} *and* 13_{th}) from the 52 SNP markers (Sanchez et al. 2006a) have been developed and validated using the polymer POP-4™ and the ABI PRISM® 310. As mentioned in the original paper (Sanchez et al. 2006a), the ABI PRISM®310 has lower sensitivity than the ABI PRISM® 3100. Most of their studies were carried out using ABI PRISM® 3100 or ABI 3100-Avant Genetic Analyzers. As for this study, the capability of a 3500 in detecting designated SNP assays was assessed as this instrument will be used to analyze a number of Malaysian population samples and forensic real-casework samples.

4.4.1. Corroboration analysis with the 3500

Any instrumentation used for forensic applications must be accurate and allow correct allelic designation of samples processed (Koumi et al. 2004). Therefore, for comparison with the ABI PRISM® 310, the same multiplex assays and positive controls generated from the previous study (*Chapter 3, Part two*), were analyzed using POP-7™ and a 3500 genetic analyzer. However, an unexpected scenario had occurred where the results demonstrated in this study show the resolution of the alleles on POP-7™ was very poor.

4.4.2. DNA separation and resolution analysis with different polymer types

To more closely examine the problem encountered, further analysis was performed using different parameters and polymer types as mentioned in *section 4.2.2*. This experiment has been repeated again with different batches of capillary arrays, buffers and polymers, and using recommended run voltage and temperature. Similar results were obtained. Results show consistently that there are significant mobility differences between alleles analyzed on POP-7™ with the actual SNP allele sizes for all Control DNA and tested samples. Besides that, polymer POP-6™ was also found to be the most

suitable polymer to be used for 13-plex assays detection on the 3500. It gave better resolution than other type of polymers, even for sizes as small as 20 nt.

Bekeart and co-workers have experienced the unresolved alleles at size ranging from 30 to 40 nt of 34-SNPplex assay when using POP-7™ on ABI 3130xl genetic analyzer, which was suggested to be due to the higher viscosity of the polymer (Bekaert et al. 2011).

According to Wenz, a sample with a consistent conformation between fragments could be an indicator of a resolution for a CE system. He also added that for single-nucleotide genotyping, it requires the separation of length variant alleles from common alleles that can differ in size by one nucleotide. In addition, Wenz et al. has suggested to use a higher concentrated polymer formulation (POP-6™) for DNA sequencing, because it provides 1-nucleotide separation up to at least 400 nucleotides, although POP-4 is sufficient for most genotyping experiments, but it provides 1-nucleotide separation beyond 250 nucleotides in some instances (Wenz et al. 1998).

Apart from the resolution problem, which occurred within the range of 20 to 60 nt, all peaks height, size and shape were good and no other artifacts were encountered; such as shoulder peaks or blobs, in any profiles analyzed with the different types of polymers on both genetic analyzers. Therefore, no further analysis was carried out.

4.4.3. Overview of polymer solutions, includes POP-4™, POP-6™ and POP-7™

There are several factors affecting the resolution in a CE system, which include capillary length, polymer concentrations, the buffer’s ionic strength and separation voltage (Butler 2011). But polymer solution is one of the most important parameters in CE since it determines the migration behavior, including the resolution of charged molecule.

Many studies had been carried out in finding, developing and improving the suitable polymer to be used in separating DNA efficiently (Wenz 1994, Rosenblum et al. 1997,

Madabhushi et al. 1997, Mandabhushi 1998, Lazaruk et al. 1998, Wenz et al. 1998, Behr et al. 1999, Madabhushi 2001, Albarghouthi et al. 2002, Fredlake et al. 2008, Maruyama 2005). Most of them have been working on the development of low viscosity matrices to allow the refilling of the capillary after each analysis. A 'replaceable' or 'replenishable' polymer system extends the life time of a capillary, prevents polymer contamination, and avoids sample carry-over.

Based on theoretical considerations and proved by experiments, Madabhushi and collegues at Applied Biosystems stated that pDMA is close to an ideal matrix for DNA separation; it can also be pumped through the capillaries in less than a minute and provides a very high resolution for the separation of dsDNA as well as ssDNA (Madabhushi 1998). The most important property of pDMA is its self-coating ability: the hydrophilic pDMA chains adhere to the capillary wall, therefore effectively suppressing electroosmotic flow (Behr et al. 1999). The polymer solution needs to be simply flushed through the capillaries to coat them. Alternatively, addition of small amounts of these coating polymers to other sieving polymers allows the use of uncoated capillaries (Madabhushi 1998). Using the 3500 over 1000 runs has been possible without any decline in capillary performance.

Poly (dimethyl acrylamide)/ (PDMA) and polyvinylpyrrolidone (PVP) are non-covalent coating polymers and were used for DNA sequencing. These polymers suppress EOF and DNA-wall interactions better than PEO due to their strong adsorption characteristics. At least 100 sequencing runs were performed in a PDMA coated capillary before surface regeneration was required (Madabhushi 1998, Madabhushi 2001).

The composition of the POP-4™ and POP-6™ were already known (Wenz 1994, Madabhushi et al. 1997, Madabhushi 1998) , but for POP-7™, only information about its characteristic and analysis features are acknowledged, such as POP-7™ was designed for a wide range of short-read to long-read sequencing applications. However, there was one chemical analysis of POP7 that had been conducted by Maruyama to study the main ingredient of POP-7, and from the results obtained, he has suggested

that POP-7™ consists mainly of Linear polyacrylamide (LPA) and trace amount of pDMA, urea and N-tris(hydroxymethyl)methyl-3-aminopropanesulfonic acid (TAPS) (Maruyama 2005).

In the forensic field, POP-7™ was the newest polymer developed specifically to be used with ABI 3700 series genetic analyzer. Kazim et al. reported that with POP-7™, the occurrence of the artifacts such as, spikes and stutters were greatly reduced (Detwiler et al. 2004). POP-7™ also was introduced to be used with the flexible 3130 Series Systems, to perform sequencing and fragment analysis (Applied Biosystems 2004).

Sizing differences resulting from mobility changes between various types of polymer are more apparent for sequences smaller than 50 nt. Due to the nature of the polymer, smaller fragments (< 50 nt) run on POP-7™ polymer on the 3130/*xl* Genetic Analyzers and on the 3730/*xl* DNA Analyzers may have slightly different mobility. Applied Biosystems has recommended that genotyping projects should be started and completed on the same instrument, using consistent run conditions. Besides that, the single-base extension primers should be designed not shorter than 25 nt to prevent sizing inconsistency. In addition, the spacing between the alleles should be at least 6-8 nt for primers that are less than 30 nt and 4-6 nt for primers that are greater than 30 nt (Applied Biosystems 2010b).

4.4.4. Verification studies under optimal conditions and parameters on the 3500 Genetic Analyzer

Koumi and colleagues described the need to evaluate and validate the platforms for forensic use, where a number of key acceptance criteria were examined, including correct designation of known sample profiles, precision, resolution, and sensitivity (Koumi et al. 2004).

In this study, all criteria were assessed to ensure the reliability of the assays and the genetic analyzer systems. Although no allelic ladder was used, the whole process was evaluated using the TaqMan DNA Controls and control DNA 9947A.

Koumi and colleagues also stated that the instrument's precision study, consisting of a number of allelic ladder runs or demonstrated an internal size standard migrating at a linear rate in achieving precise sizing (Koumi et al. 2004). Sizing precision is an indicator that the instrument is performing consistently rather than sizing accuracy, which is a measure of how closely fragments are sized to their true length. Differences are observed in assigned allele sizes when using different instruments, separation matrices, and size standards (Lazaruk et al. 1998, Vainer et al. 1997).

In this study, some additional tests were conducted on the polymers to ensure how accurate 3500 in resolving alleles, where the difference in length (nt) were measured between the alleles electrophoresed on the polymer POP-4™, POP-6™ and POP-7™. There were significant deviations between the alleles electrophoresed on POP-7™ compared to the other two polymers.

Therefore, POP-6™ was chosen to be used for SNP genotyping of population and casework samples. A few other parameters such as the sensitivity of DNA level detection for 13-plex assays and others also had been evaluated before performing the 13-plex assays on the samples.

From the evaluation, it shows that although the samples loaded on the 3500 was as low as 0.4 µl (of sample following SBE reaction and clean-up) full SNP profile were generated with a threshold of approximately 3000 RFU. Early in the development of DNA testing with CE, it was demonstrated that a simple dilution of the sample in water or deionized formamide can be an effective method for sample preparation because the sample ionic strength is reduced relative to the buffer ionic strength (Behr et al. 1999).

In addition, the SBE reaction mix has been also reduced from the original quantity, because the 3500 was found to be more sensitive in detecting the 13-plex assays compared to the ABI 310 genetic analyzer.

The results also showed consistency in triplicates that the 13-plex assays can be detected for DNA input as low as 30 pg, although in some instances, a few of the

population samples have one or two allele(s) drop-out. A drop out observed at 30 pg template DNA probably due to variations introduced during the PCR set-up procedure rather than instrument effects. The best SNP profile obtained was at input DNA 60 pg and above. Koumi and colleagues suggested that sensitivity was examined using dilutions of PCR products and measured by the ability to generate a full profile by the instruments (Koumi et al. 2004).

Finally, the 13-plex assays were arranged as a full SNP profile (52 markers) as four different panels (eg. Figure 4-12 and 4-13). In this way it is a lot easier to interpret the SNP data.

CHAPTER 5

SNP GENOTYPING OF MALAYSIAN POPULATION SAMPLES WITH 13-PLEX PCR/SBE ASSAYS

SNP ANALYSIS OF MALAYSIAN POPULATION SAMPLES

WITH 13-PLEX PCR/SBE ASSAYS (13_{st}, 13_{nd}, 13_{rd} and 13_{th})

5.1. INTRODUCTION

SNPs could be polymorphic in one population but completely monomorphic in another population (Pfaff et al. 2004, Sanchez et al. 2006a), while others are known to be polymorphic in all major population groups. For human identification purposes, loci with low F_{ST} values and a heterozygosity of 40% are generally considered adequate However, a major problem with SNPs is that the frequency of an allele can range from zero to one among divergent populations; causing a very large dependence of the match probability on the population allele frequencies used for the calculation, potentially to a level many fold greater than for CODIS markers (Kidd et al. 2006).

Before any new forensic markers are used, it is very important to generate appropriate allele frequency databases that can be referred to. This is especially important in markers such as SNPs, where allele frequencies can vary considerably between populations. Information on allele frequency measurements from a group or representative sample set (usually 100 individuals), is used. These individuals are completely anonymous and only classified with their ethnicity (Butler 2011). With modern cosmopolitan populations, allele frequencies for all forensically relevant populations should be known to avoid courtroom challenges (Kidd et al. 2006, Pakstis et al. 2007, John M. Butler, Michael D. Coble, Peter M. Vallone 2007).

In Malaysia, there is no accessible intelligence DNA databank system yet such as NDNAD (UK) or CODIS (USA). However there are published allele frequency data for 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.

5.1.1. 52-SNPforID plex

Sanchez and colleagues have studied these populations (Table 5-1) in order to study the chosen SNP markers to be used in forensic applications. They have reported quite huge SNP population data and also carried out studies on the forensic samples, which includes kinship study.

Table 5-1: The populations included in the Sanchez study (Sanchez et al. 2006a).

Population	Total number of populations analyzed with 52-plex SNPforID
European:	
Danish	156
German	49
Turkish	96
Asians:	
Chinese	63
Taiwanese	43
Japanese	7
Thais	33
Greenlander	149
African:	
Somali	104
Total	**700**

5.1.2. SNPforID Browser

Several members of the European Forensic community launched a project in 2003 known as SNP*for*ID. This group has developed large multiplexed SNP assays for human identification that comprised 52 SNP markers (Sanchez et al. 2006a, Butler 2011). This group has also established a web-based tool, termed as "SNP*for*ID Browser" (shown in Figure 5-1) for the visualization or query of SNP allele frequency data and other population parameters generated by the consortium (Amigo et al. 2008a).

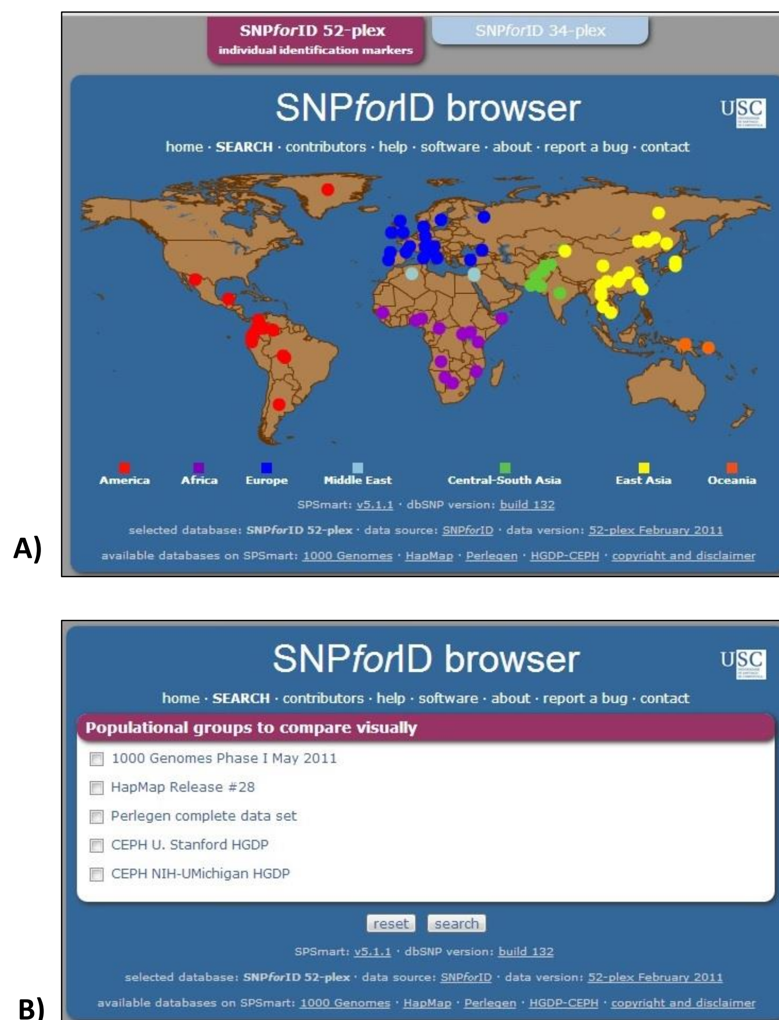


Figure 5-1: (A) The SNPforID Browser homepage, and (B) several population databases that can be used for data comparison. [Taken from (SNPforID Consortium , Amigo et al. 2008a).

5.2. RATIONALE OF THE STUDY

In this study, 52 SNP markers have been chosen to be applied to Malaysian forensic casework samples. Hence, before applying these SNP markers to the samples, various forensic genetic and population genetic parameters needed to be collected.

5.2.1. Malaysian population samples

The population samples (Lim et al. 2001, Seah et al. 2003, Izuan et al. 2005, Chang et al. 2007) were genotyped with the four 13-plex assays *(13_{st}, 13_{nd}, 13_{rd} and 13_{th})* designed in this study (*Chapter 3, Part two*) using optimal analysis parameters (as mentioned in Chapter 2, section 2.3 and 2.7).

5.2.2. Data Collection and statistical evaluation

The population samples from three ethnic groups were analyzed with GeneMapper *ID-X® (*Applied Biosystems). Forensic parameters such as allele frequencies, power of discrimination (PD), random match probability (MP), paternity index (PI), probability of exclusion (PE) and polymorphism information content (PIC) were determined using Powerstats software (Jones 1972, Botstein et al. 1980). Observed and expected heterozygosity estimates, exact tests for Hardy-Weinberg equilibrium (HWE) and linkage equilibrium (LE), and pairwise F_{ST} values were calculated using Arlequin version 3.5.1.3 software (Excoffier, Lischer 2010).

5.2.3. The Snipper app suite online software

Phillips and colleagues (2007) have established a straightforward Bayesian system for predicting ancestral origin and to estimate misclassification rates by testing the CEPH human genome diversity cell line panel (CEPH-HGDP) comprising samples of confirmed

geographic origin. This classification algorithm has been incorporated into an open access web portal to allow simple analysis of SNP profiles, including those with partial data (Phillips et al. 2007b). This portal was enhanced to allow analysis of a user's custom population(s) and SNP markers (Pereira et al. 2012, Fondevila et al. 2013b) with the same Bayesian classification algorithm and error estimation systems. Training sets for the classification algorithm were created for each population-group by combining two population samples comprising: sub-Saharan Africans (60 Mozambican and 60 Somali), Europeans (60 Galician from Spain and 60 Danish) and East Asians (60 Mainland Chinese and 60 Taiwanese). The CEPH-HGDP panel comprising 1064 samples from 51 geographically diverse populations was also used to test classification performance with new profiles.

This software was used in this study to demonstrate the performance of the 52 SNP markers in inferring the ancestral origin of a DNA sample. 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. 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 (Snipper SNP-Indel application portal).

5.2.4. Data comparison between Malaysia and neighbouring countries

SNP*for*ID Browser has genotypes and statistical information of several sets of population groups (Figure 5-2) that have been analyzed with the SNP*for*ID 52-plex (Sanchez et al. 2006a); and includes Africa, America, Europe, Middle East, Central South-Asia, East Asia and Oceania (SNPforID Consortium).

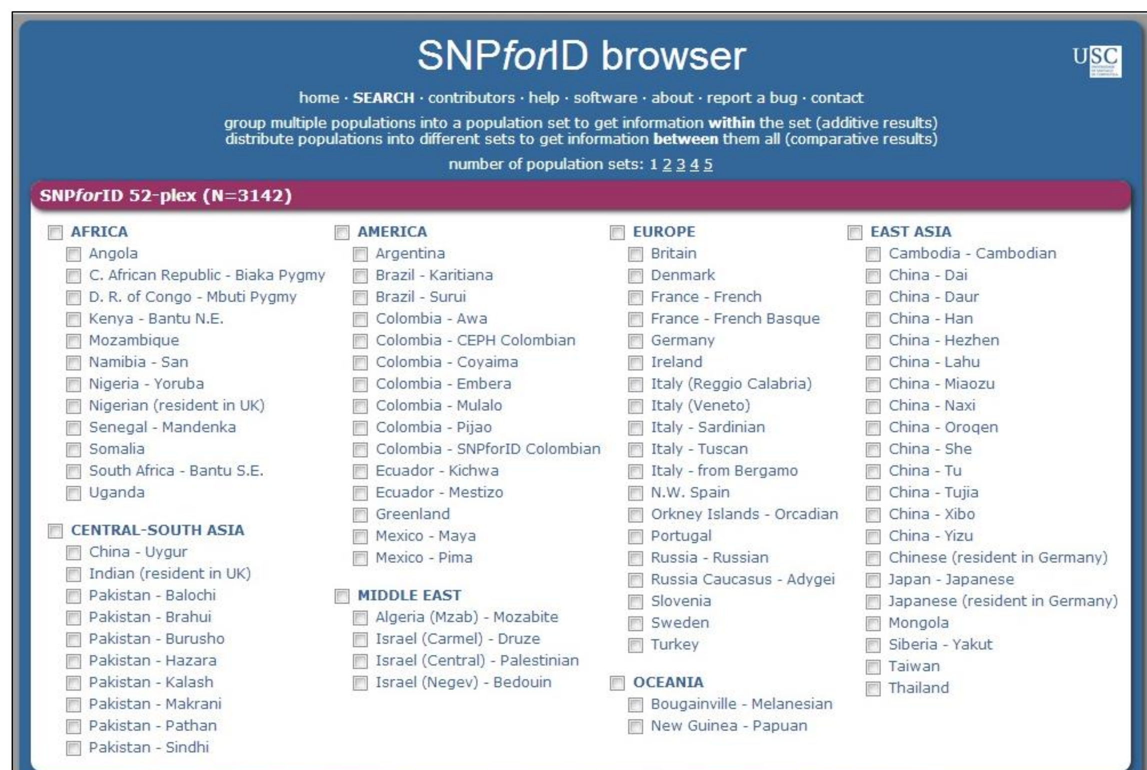


Figure 5-2: Several groups of populations that involved in SNP*for*ID project [Taken from (SNPforID Consortium)].

Data generated from other neighbouring countries as well as from nearby regions such as East Asia, Central-South Asia and Oceania by the SNPforID Consortium (Table 5-2) were used for comparison with data collected from Malaysian population samples. Data from Europe were also used as a reference in this study, as these 52 SNP markers were originally designed and selected based on the maximum heterozygosity found in European populations.

Table 5-2: Groups of populations used for comparison in this study from the (SNPforID Consortium).

SNP*for*ID 52-plex									
Malaysia (4 x 13-plex assays) (N=325)	N	East Asia (N=378)	N	Central-South Asia (N=234)	N	Oceania (N=28)	N	Europe (N=1140)	N
Malays	109	Cambodia-Cambodian	10	China Uygur	10	Bougainville -Melanesian	11	Britain	18
Chinese	107	China-Dai	10	Indian (in UK)	32	New Guinea Papuan	17	Denmark	156
Indians	109	China-Daur	10	Pakistan-Balochi	24			France-French	28
		China-Han	44	Pakistan-Brahui	25			France-French Basque	24
		China-Hezhen	9	Pakistan-Burusho	25			Germany	49
		China-Lahu	8	Pakistan-Hazara	22			Ireland	24
		China-Miaou	10	Pakistan-Kalash	23			Italy (Reggio Calabria)	33
		China-Naxi	9	Pakistan-Makrani	25			Italy (Veneto)	30
		China-Oroqen	9	Pakistan-Pathan	24			Italy -Sardinian	28
		China-She	10	Pakistan-Sindhi	24			Italy-Tuscan	8
		China-Tu	10					Italy-Bergamo	13
		China-Tujia	10					N.W. Spain	151
		China-Xibo	9					Orkney Islands-Orcadian	15
		China-Yizu	10					Portugal	193
		Chinese (in Germany)	63					Russia-Russian	25
		Japan-Japanese	29					Russia Caucasus-Adygei	17
		Japanese (in Germany)	7					Slovenia	154
		Mongola	10					Sweden	78
		Siberia-Yakut	25					Turkey	96
		Taiwan	43						
		Thailand	33						

5.3. RESULTS

5.3.1. Allele Frequencies

In this study, a total of 325 individuals comprising 109 Malays, 107 Chinese and 109 Indians were genotyped using 52 SNP markers. Table 5-3, shows data comparison of allele frequencies obtained in these three ethnic groups with the reference data from other countries for comparison (SNPforID Consortium).

The minimum allele frequency observed in Malaysian population samples is 0.079 in the Chinese and 0.096 in the Malays (both at SNP code marker 29), and 0.142 in the Indians (at SNP code marker 16). The values were italicized and emboldened. No values below 0.023 (the calculated minimum allele frequency of 5/2 N) were observed.

It is notable that allele frequencies of most SNP markers in Malays and Chinese are very similar to each other and also to the East Asian populations. Whilst for the Indians, the allele frequencies observed for most SNP markers are very similar to the Central South Asia and European populations.

Table 5-3: Allele frequencies data of Malaysian and neighbouring countries populations

Individual Identification SNPs (IISNPs)			Allele frequencies						
Marker Code	SNP	Allele	Malaysia			East Asia (N = 378)	Central - South Asia (N = 234)	Oceania (N = 28)	Europe (N = 1140)
			Malay (N=109)	Chinese (N=107)	Indian (N=109)				
1	**rs1490413**	C	0.427	0.444	0.486	0.423	0.482	0.333	0.564
		T	0.573	0.556	0.514	0.577	0.518	0.667	0.436
2	**rs876724**	C	0.647	0.682	0.674	0.564	0.724	0.981	0.685
		T	0.353	0.318	0.326	0.436	0.276	0.019	0.315
3	**rs1357617**	A	0.161	0.178	0.275	0.188	0.293	0.054	0.293
		T	0.839	0.822	0.725	0.812	0.707	0.946	0.707
4	**rs2046361**	A	0.454	0.449	0.385	0.417	0.432	0.593	0.350
		T	0.346	0.551	0.615	0.583	0.568	0.407	0.650
5	**rs717302**	A	0.734	0.855	0.601	0.848	0.625	0.696	0.503
		G	0.266	0.145	0.399	0.152	0.375	0.304	0.497
6	**rs1029047**	A	0.339	0.290	0.495	0.357	0.370	0.240	0.396
		T	0.661	0.710	0.505	0.643	0.630	0.760	0.604
7	**rs917118**	A	0.225	0.285	0.413	0.267	0.448	0.232	0.283
		G	0.775	0.715	0.587	0.733	0.552	0.768	0.717
8	**rs763869**	C	0.431	0.421	0.216	0.346	0.287	0.429	0.486
		T	0.569	0.579	0.784	0.654	0.713	0.571	0.514
9	**rs1015250**	C	0.573	0.491	0.596	0.535	0.716	0.179	0.786
		G	0.427	0.509	0.404	0.465	0.284	0.821	0.214
10	**rs735155**	C	0.179	0.126	0.339	0.190	0.393	0.643	0.536
		T	0.821	0.874	0.661	0.810	0.607	0.357	0.464
11	**rs901398**	C	0.257	0.294	0.335	0.265	0.274	0.214	0.326
		T	0.743	0.706	0.665	0.735	0.726	0.786	0.674
12	**rs2107612**	A	0.706	0.738	0.500	0.871	0.728	0.411	0.673
		G	0.294	0.262	0.500	0.129	0.272	0.589	0.327
13	**rs1886510**	A	0.142	0.159	0.271	0.117	0.209	0.161	0.436
		G	0.858	0.841	0.729	0.869	0.700	0.804	0.505
14	**rs1454361**	A	0.564	0.570	0.615	0.492	0.476	0.143	0.531
		T	0.436	0.430	0.385	0.508	0.524	0.857	0.469
15	**rs2016276**	C	0.349	0.350	0.261	0.377	0.214	0.054	0.231
		T	0.651	0.650	0.739	0.623	0.786	0.946	0.769
16	**rs729172**	G	0.844	0.860	0.858	0.836	0.726	0.946	0.598
		T	0.156	0.140	***0.142***	0.164	0.274	0.054	0.402
17	**rs740910**	A	0.867	0.916	0.839	0.922	0.738	0.815	0.702
		G	0.133	0.084	0.161	0.078	0.262	0.185	0.298
18	**rs1493232**	G	0.601	0.631	0.390	0.511	0.242	0.714	0.336
		T	0.399	0.369	0.610	0.489	0.758	0.286	0.664
19	**rs719366**	C	0.335	0.215	0.312	0.231	0.377	0.778	0.373
		T	0.665	0.785	0.688	0.769	0.623	0.222	0.627
20	**rs1031825**	G	0.463	0.486	0.546	0.568	0.577	0.714	0.731
		T	0.537	0.514	0.454	0.432	0.423	0.286	0.269
21	**rs722098**	A	0.583	0.486	0.555	0.446	0.680	0.500	0.782
		G	0.417	0.514	0.445	0.554	0.320	0.500	0.218
22	**rs733164**	A	0.280	0.164	0.183	0.145	0.252	0.463	0.296
		G	0.720	0.836	0.817	0.855	0.748	0.537	0.704
23	**rs826472**	A	0.202	0.121	0.390	0.174	0.312	0.648	0.376
		G	0.798	0.879	0.610	0.826	0.688	0.352	0.624
24	**rs2831700**	C	0.427	0.542	0.225	0.493	0.335	0.519	0.408
		T	0.573	0.458	0.775	0.507	0.665	0.481	0.592
25	**rs873196**	A	0.839	0.911	0.812	0.864	0.703	0.679	0.599
		G	0.161	0.089	0.188	0.136	0.297	0.321	0.401
26	**rs1382387**	G	0.413	0.393	0.408	0.304	0.293	0.185	0.319
		T	0.587	0.607	0.592	0.696	0.707	0.815	0.681

Table 5-3: Allele frequencies data of Malaysian and neighbouring countries populations (continued).

Individual Identification SNPs (IISNPs)			Allele frequencies						
Marker Code	SNP	Allele	Malaysia			East Asia (N = 378)	Central - South Asia (N = 234)	Oceania (N = 28)	Europe (N = 1140)
			Malay (N=109)	Chinese (N=107)	Indian (N=109)				
27	rs2111980	A	0.610	0.603	0.509	0.619	0.491	0.648	0.456
		G	0.390	0.397	0.491	0.381	0.509	0.352	0.544
28	rs2056277	A	0.170	0.121	0.206	0.133	0.167	0.056	0.259
		G	0.830	0.879	0.794	0.867	0.833	0.944	0.741
29	rs1024116	A	***0.096***	***0.079***	0.284	0.121	0.412	0.500	0.573
		G	0.904	0.921	0.716	0.879	0.588	0.500	0.427
30	rs727811	A	0.541	0.640	0.674	0.688	0.513	0.661	0.552
		C	0.459	0.360	0.326	0.312	0.487	0.339	0.448
32	rs1413212	C	0.422	0.341	0.390	0.545	0.656	0.574	0.697
		T	0.578	0.659	0.610	0.455	0.344	0.426	0.303
33	rs938283	C	0.284	0.150	0.161	0.150	0.103	0.019	0.169
		T	0.716	0.850	0.839	0.850	0.897	0.981	0.831
34	rs1979255	C	0.459	0.444	0.372	0.436	0.435	0.268	0.339
		G	0.541	0.556	0.628	0.564	0.565	0.732	0.661
35	rs1463729	A	0.495	0.533	0.454	0.449	0.507	0.058	0.556
		G	0.505	0.467	0.546	0.551	0.493	0.942	0.444
36	rs2076848	A	0.252	0.322	0.436	0.340	0.446	0.286	0.441
		T	0.748	0.678	0.564	0.660	0.554	0.714	0.559
37	rs1355366	C	0.330	0.164	0.518	0.157	0.473	0.096	0.421
		T	0.670	0.836	0.482	0.843	0.527	0.904	0.579
38	rs907100	C	0.468	0.533	0.463	0.550	0.550	0.786	0.606
		G	0.532	0.467	0.537	0.450	0.450	0.214	0.394
39	rs354439	A	0.408	0.421	0.569	0.431	0.638	0.875	0.418
		T	0.592	0.579	0.431	0.569	0.362	0.125	0.582
40	rs2040411	A	0.326	0.332	0.472	0.271	0.461	0.750	0.629
		G	0.674	0.668	0.528	0.729	0.539	0.250	0.371
41	rs737681	C	0.835	0.864	0.606	0.827	0.651	0.982	0.594
		T	0.165	0.136	0.394	0.173	0.349	0.018	0.406
42	rs2830795	A	0.399	0.537	0.706	0.516	0.770	0.821	0.709
		G	0.601	0.463	0.294	0.484	0.230	0.179	0.291
43	rs251934	A	0.867	0.888	0.688	0.875	0.758	0.696	0.614
		G	0.133	0.112	0.312	0.125	0.242	0.304	0.386
44	rs914165	C	0.706	0.640	0.578	0.689	0.571	0.286	0.604
		T	0.294	0.360	0.422	0.311	0.429	0.714	0.396
45	rs10495407	C	0.720	0.696	0.844	0.671	0.740	0.978	0.643
		T	0.280	0.304	0.156	0.329	0.260	0.022	0.357
46	rs1360288	A	0.220	0.350	0.486	0.347	0.373	0.074	0.335
		G	0.780	0.650	0.514	0.653	0.627	0.926	0.665
48	rs964681	C	0.335	0.360	0.436	0.298	0.423	0.375	0.417
		T	0.665	0.640	0.564	0.702	0.577	0.625	0.583
49	rs1005533	A	0.197	0.290	0.335	0.332	0.399	0.519	0.536
		G	0.803	0.710	0.665	0.668	0.601	0.481	0.464
50	rs8037429	A	0.422	0.374	0.390	0.456	0.543	0.583	0.516
		G	0.578	0.626	0.610	0.544	0.457	0.417	0.484
51	rs891700	A	0.472	0.458	0.422	0.504	0.417	0.357	0.475
		G	0.528	0.542	0.578	0.496	0.583	0.643	0.525
52	rs1335873	A	0.716	0.617	0.702	0.658	0.728	0.923	0.691
		T	0.284	0.383	0.298	0.342	0.272	0.077	0.309
53	rs1028528	A	0.555	0.589	0.550	0.688	0.709	0.232	0.742
		G	0.445	0.411	0.450	0.312	0.291	0.768	0.258
54	rs1528460	A	0.528	0.477	0.523	0.616	0.603	0.250	0.690
		G	0.472	0.523	0.477	0.384	0.397	0.750	0.310

*Bold and italic number(s) represent minor allele frequency in three ethnic groups: the Malays, Chinese and Indians.

5.3.2. Tests for Hardy-Weinberg Equilibrium (HWE)

The observed and expected heterozygosity values among the Malaysian ethnic groups are summarized in Table 5-4. The highlighted cells in the P-value column show significant deviation from HWE at p<0.05 using an exact test, and in H_o column, shows minimum observed heterozygosity in each ethnic group.

As shown, the minimum observed heterozygosities (highlighted in yellow) for Malays, Chinese and Indians are 0.16, 0.15 and 0.17, respectively and the minimum expected heterozygosities (highlighted in blue) are 0.17, 0.15 and 0.23, respectively. The average observed heterozygosities for Malays, Chinese and Indians are 0.42, 0.41 and 0.46, respectively and the average expected heterozygosities are 0.41, 0.40 and 0.44.

Significant departures from HWE, at p values <0.05 were observed at 3 SNP markers in Malays, 4 SNP markers in Chinese and 9 SNP markers in Indian samples as shown in Table 5-4. The rs2107612 marker (*code 12*), rs722098 marker (*code 21*), rs2076848 marker (*code 36*) and rs907100 marker (*code 38*) in the Indians, and rs1528460 marker (*code 54*) in the Chinese and also in the Indians, showed the lowest p value, that is p=0.0.

As for SNP markers that show p values <0.05, the Bonferroni correction at p <0.00096 (0.05 divided by 52 i.e. the number of loci tested) was applied and data were re-analyzed. After the Bonferroni correction, significant deviation(s) from HWE was observed at 1 SNP marker (code marker 26) in the Malays, 2 SNP markers (code marker 46 and 54) in the Chinese and 5 SNP markers (code marker 12, 21, 36, 38 and 54) in the Indians.

Observed and expected heterozygosity values of Malaysian population were then compared with data collated from the SNP*for*ID Browser as demonstrated in Table 5-5. The values that represented HWE departures in Malaysian population are highlighted in red, and the corresponding values from this locus are emboldened and italicized for the other populations but found to be in HWE in these populations.

Table 5-4: Observed (H_o) and expected [H_e] heterozygosities and P values from an exact test for HWE across 52 SNPs typed in 325 individuals from three ethnic groups. Yellow represents minimum H_o; blue represents minimum H_e; and red represents p value <0.05.

Marker Code	SNP (rs#)	Malay			Chinese			Indian		
		Obs.$_H$	Exp.$_H$	P-value	Obs.$_H$	Exp.$_H$	P-value	Obs.$_H$	Exp.$_H$	P-value
1	rs1490413	0.50	0.50	1.00	0.41	0.49	0.12	0.51	0.50	0.85
2	rs876724	0.50	0.46	0.40	0.40	0.43	0.51	0.41	0.44	0.52
3	rs1357617	0.28	0.27	0.73	0.29	0.29	1.00	0.50	0.40	0.0166
4	rs2046361	0.50	0.50	1.00	0.54	0.50	0.43	0.44	0.48	0.54
5	rs717302	0.33	0.39	0.14	0.21	0.25	0.23	0.55	0.48	0.17
6	rs1029047	0.40	0.44	0.39	0.42	0.42	1.00	0.46	0.50	0.44
7	rs917118	0.32	0.35	0.41	0.48	0.41	0.10	0.55	0.49	0.23
8	rs763869	0.42	0.49	0.17	0.47	0.49	0.70	0.39	0.34	0.15
9	rs1015250	0.47	0.49	0.70	0.48	0.50	0.70	0.44	0.48	0.43
10	rs735155	0.28	0.30	0.35	0.23	0.22	1.00	0.42	0.45	0.53
11	rs901398	0.38	0.38	1.00	0.35	0.42	0.10	0.43	0.45	0.83
12	rs2107612	0.38	0.44	0.13	0.44	0.38	0.20	0.72	0.50	*0.00000
13	rs1886510	0.28	0.25	0.12	0.32	0.27	0.07	0.43	0.40	0.47
14	rs1454361	0.47	0.49	0.70	0.54	0.49	0.33	0.46	0.48	0.84
15	rs2016276	0.42	0.46	0.53	0.44	0.46	0.68	0.31	0.38	0.08
16	rs729172	0.25	0.26	0.72	0.22	0.24	0.43	0.17	0.23	0.0171
17	rs740910	0.27	0.23	0.21	0.17	0.16	1.00	0.25	0.27	0.47
18	rs1493232	0.52	0.48	0.43	0.51	0.47	0.40	0.44	0.48	0.43
19	rs719366	0.49	0.45	0.40	0.34	0.34	1.00	0.42	0.43	0.83
20	rs1031825	0.52	0.50	0.70	0.54	0.50	0.44	0.56	0.50	0.25
21	rs722098	0.49	0.49	1.00	0.54	0.50	0.44	0.76	0.50	*0.00000
22	rs733164	0.39	0.40	0.81	0.23	0.27	0.15	0.31	0.30	1.00
23	rs826472	0.35	0.32	0.56	0.22	0.21	1.00	0.45	0.48	0.55
24	rs2831700	0.49	0.49	1.00	0.54	0.50	0.44	0.35	0.35	1.00
25	rs873196	0.23	0.27	0.15	0.15	0.16	0.59	0.32	0.31	0.76
26	rs1382387	0.68	0.49	*0.00000	0.58	0.48	0.05	0.58	0.49	0.05
27	rs2111980	0.52	0.48	0.42	0.46	0.48	0.69	0.54	0.50	0.45
28	rs2056277	0.32	0.28	0.30	0.19	0.21	0.18	0.34	0.33	1.00
29	rs1024116	0.16	0.17	0.25	0.15	0.15	1.00	0.33	0.41	0.05
30	rs727811	0.51	0.50	0.85	0.48	0.46	0.83	0.54	0.44	0.0194
32	rs1413212	0.61	0.49	0.0194	0.47	0.45	0.83	0.61	0.48	0.0025
33	rs938283	0.40	0.42	0.82	0.22	0.26	0.25	0.25	0.27	0.47
34	rs1979255	0.44	0.50	0.25	0.48	0.50	0.70	0.43	0.47	0.42
35	rs1463729	0.51	0.50	0.85	0.49	0.50	0.85	0.56	0.50	0.25
36	rs2076848	0.43	0.39	0.32	0.46	0.44	0.82	0.71	0.49	*0.00000
37	rs1355366	0.51	0.44	0.13	0.33	0.27	0.07	0.52	0.50	0.70
38	rs907100	0.35	0.50	0.0016	0.43	0.50	0.17	0.27	0.50	*0.00000
39	rs354439	0.43	0.49	0.32	0.52	0.49	0.55	0.46	0.49	0.56
40	rs2040411	0.45	0.45	1.00	0.42	0.45	0.66	0.56	0.50	0.25
41	rs737681	0.29	0.28	0.73	0.25	0.23	0.69	0.42	0.48	0.23
42	rs2830795	0.50	0.48	0.70	0.59	0.50	0.08	0.44	0.42	0.65
43	rs251934	0.21	0.23	0.40	0.19	0.20	0.62	0.46	0.43	0.66
44	rs914165	0.35	0.42	0.11	0.42	0.46	0.40	0.46	0.49	0.56
45	rs10495407	0.39	0.41	0.64	0.42	0.42	1.00	0.31	0.26	0.07
46	rs1360288	0.33	0.35	0.78	0.65	0.46	*0.00000	0.57	0.50	0.18
48	rs964681	0.45	0.45	1.00	0.51	0.47	0.40	0.49	0.49	1.00
49	rs1005533	0.32	0.32	1.00	0.45	0.41	0.48	0.50	0.45	0.20
50	rs8037429	0.51	0.49	0.70	0.55	0.47	0.10	0.43	0.48	0.32
51	rs891700	0.50	0.50	1.00	0.43	0.50	0.17	0.54	0.49	0.32
52	rs1335873	0.44	0.41	0.49	0.50	0.47	0.55	0.38	0.42	0.36
53	rs1028528	0.58	0.50	0.12	0.39	0.49	0.0443	0.46	0.50	0.44
54	rs1528460	0.52	0.50	0.70	0.80	0.50	*0.00000	0.70	0.50	*0.00000
Mean		0.42	0.41		0.41	0.40		0.46	0.44	
s.d.		0.11	0.09		0.14	0.12		0.12	0.08	

Table 5-5: Comparison of Observed (H_o) and Expected (H_e) heterozygosities obtained in this study with the reference data (SNPforID Consortium). Red represents p value <0.05.

Individual Identification SNPs (IISNPs)			Malaysia (N = 325)						East Asia (N = 378)		Central-South Asia (N = 234)		Oceania (N = 28)		Europe (N = 1140)	
			Malay (N = 109)		Chinese (N = 107)		Indian (N = 109)									
Marker Code	SNP (rs#)	Alleles detected	Ho	He	Ho	He	Ho	He	Ho	He	Ho	He	Ho	He	Ho	He
1	rs1490413	A/G	0.50	0.50	0.41	0.49	0.51	0.50	0.47	0.49	0.46	0.50	0.44	0.44	0.48	0.49
2	rs876724	C/T	0.50	0.46	0.40	0.43	0.41	0.44	0.44	0.49	0.38	0.40	0.04	0.04	0.44	0.43
3	rs1357617	A/T	0.28	0.27	0.29	0.29	0.50	0.40	0.31	0.31	0.46	0.41	0.11	0.10	0.39	0.41
4	rs2046361	A/T	0.50	0.50	0.54	0.50	0.44	0.48	0.46	0.49	0.52	0.49	0.22	0.48	0.47	0.46
5	rs717302	A/G	0.33	0.39	0.21	0.25	0.55	0.48	0.23	0.26	0.47	0.47	0.46	0.42	0.46	0.50
6	rs1029047	A/T	0.40	0.44	0.42	0.42	0.46	0.50	0.47	0.50	0.50	0.47	0.40	0.37	0.48	0.48
7	rs917118	C/T	0.32	0.35	0.48	0.41	0.55	0.49	0.40	0.39	0.53	0.50	0.32	0.36	0.41	0.41
8	rs763869	C/T	0.42	0.49	0.47	0.49	0.39	0.34	0.42	0.45	0.35	0.41	0.50	0.49	0.47	0.50
9	rs1015250	C/G	0.47	0.49	0.48	0.50	0.44	0.48	0.51	0.50	0.37	0.41	0.29	0.29	0.34	0.34
10	rs735155	A/G	0.28	0.30	0.23	0.22	0.42	0.45	0.29	0.31	0.47	0.48	0.36	0.46	0.50	0.50
11	rs901398	C/T	0.38	0.38	0.35	0.42	0.43	0.45	0.45	0.39	0.39	0.40	0.29	0.34	0.45	0.44
12	rs2107612	A/G	0.38	0.44	0.44	0.38	0.72	0.50	0.20	0.23	0.38	0.40	0.32	0.48	0.44	0.44
13	rs1886510	C/T	0.28	0.25	0.32	0.27	0.44	0.40	0.22	0.23	0.41	0.46	0.32	0.33	0.46	0.55
14	rs1454361	A/T	0.47	0.49	0.54	0.49	0.46	0.48	0.50	0.50	0.41	0.50	0.21	0.25	0.49	0.49
15	rs2016276	A/G	0.42	0.46	0.44	0.46	0.31	0.38	0.42	0.47	0.33	0.34	0.11	0.10	0.34	0.36
16	rs729172	A/C	0.25	0.26	0.22	0.24	0.17	0.23	0.27	0.27	0.38	0.40	0.11	0.10	0.46	0.48
17	rs740910	A/G	0.27	0.23	0.17	0.16	0.25	0.27	0.14	0.14	0.34	0.39	0.22	0.30	0.39	0.42
18	rs1493232	C/A	0.52	0.48	0.51	0.47	0.44	0.48	0.27	0.50	0.18	0.37	0.14	0.41	0.43	0.45
19	rs719366	C/T	0.49	0.45	0.34	0.34	0.42	0.43	0.35	0.36	0.46	0.47	0.30	0.35	0.44	0.47
20	rs1031825	A/C	0.52	0.50	0.54	0.50	0.56	0.50	0.50	0.49	0.46	0.49	0.29	0.41	0.39	0.39
21	rs722098	A/G	0.49	0.49	0.54	0.50	0.76	0.50	0.47	0.49	0.36	0.44	0.31	0.50	0.33	0.34
22	rs733164	A/G	0.39	0.40	0.23	0.27	0.31	0.30	0.25	0.25	0.36	0.38	0.48	0.50	0.43	0.42
23	rs826472	C/T	0.35	0.32	0.22	0.21	0.45	0.48	0.28	0.29	0.39	0.43	0.41	0.46	0.48	0.47
24	rs2831700	G/A	0.49	0.49	0.54	0.50	0.35	0.35	0.49	0.50	0.42	0.45	0.58	0.50	0.50	0.48
25	rs873196	C/T	0.23	0.27	0.15	0.16	0.32	0.31	0.24	0.24	0.37	0.42	0.36	0.44	0.50	0.48
26	rs1382387	G/T	0.68	0.49	0.58	0.48	0.58	0.49	0.42	0.42	0.37	0.42	0.30	0.30	0.43	0.43
27	rs2111980	A/G	0.52	0.48	0.46	0.48	0.54	0.50	0.45	0.47	0.51	0.50	0.48	0.46	0.50	0.50
28	rs2056277	C/T	0.32	0.28	0.19	0.21	0.34	0.33	0.23	0.23	0.27	0.28	0.04	0.11	0.40	0.38
29	rs1024116	A/G	0.16	0.17	0.15	0.15	0.33	0.41	0.19	0.21	0.43	0.48	0.33	0.50	0.48	0.49
30	rs727811	A/C	0.51	0.50	0.48	0.46	0.54	0.44	0.44	0.43	0.55	0.50	0.54	0.45	0.50	0.50
32	rs1413212	A/G	0.61	0.49	0.47	0.45	0.61	0.48	0.48	0.50	0.41	0.45	0.48	0.49	0.41	0.42
33	rs938283	C/T	0.40	0.42	0.22	0.26	0.25	0.27	0.23	0.26	0.17	0.19	0.04	0.04	0.29	0.28
34	rs1979255	C/G	0.44	0.50	0.48	0.50	0.43	0.47	0.51	0.49	0.48	0.49	0.32	0.39	0.47	0.45
35	rs1463729	A/G	0.51	0.50	0.49	0.50	0.56	0.50	0.43	0.50	0.50	0.50	0.12	0.11	0.48	0.49
36	rs2076848	A/T	0.43	0.39	0.46	0.44	0.71	0.49	0.44	0.45	0.48	0.49	0.43	0.41	0.48	0.49
37	rs1355366	A/G	0.51	0.44	0.33	0.27	0.52	0.50	0.28	0.27	0.50	0.50	0.19	0.17	0.48	0.49
38	rs907100	C/G	0.35	0.50	0.43	0.50	0.27	0.50	0.39	0.50	0.28	0.50	0.07	0.34	0.38	0.46
39	rs354439	A/T	0.43	0.49	0.52	0.49	0.46	0.49	0.48	0.49	0.44	0.46	0.18	0.22	0.50	0.49
40	rs2040411	A/G	0.45	0.45	0.42	0.45	0.56	0.50	0.36	0.40	0.50	0.50	0.50	0.38	0.45	0.47
41	rs737681	C/T	0.29	0.28	0.25	0.23	0.42	0.48	0.28	0.29	0.40	0.45	0.04	0.04	0.46	0.48
42	rs2830795	A/G	0.50	0.48	0.59	0.50	0.44	0.42	0.51	0.50	0.33	0.35	0.29	0.29	0.44	0.41
43	rs251934	C/T	0.21	0.23	0.19	0.20	0.46	0.43	0.21	0.22	0.38	0.37	0.32	0.42	0.49	0.47
44	rs914165	A/G	0.35	0.42	0.42	0.46	0.46	0.49	0.42	0.43	0.49	0.49	0.36	0.41	0.51	0.48
45	rs10495407	G/A	0.39	0.41	0.42	0.42	0.31	0.26	0.40	0.44	0.41	0.39	0.04	0.04	0.45	0.46
46	rs1360288	C/T	0.33	0.35	0.65	0.46	0.57	0.50	0.42	0.45	0.43	0.47	0.07	0.14	0.43	0.45
48	rs964681	C/T	0.45	0.45	0.51	0.47	0.49	0.49	0.41	0.42	0.45	0.49	0.54	0.47	0.46	0.49
49	rs1005533	A/G	0.32	0.32	0.45	0.41	0.50	0.45	0.42	0.44	0.52	0.48	0.50	0.50	0.53	0.50
50	rs8037429	C/T	0.51	0.49	0.55	0.47	0.43	0.48	0.46	0.50	0.44	0.50	0.42	0.49	0.50	0.50
51	rs891700	A/G	0.50	0.50	0.43	0.50	0.54	0.49	0.47	0.50	0.48	0.49	0.43	0.46	0.48	0.50
52	rs1335873	A/T	0.44	0.41	0.50	0.47	0.38	0.42	0.44	0.45	0.36	0.40	0.15	0.14	0.44	0.43
53	rs1028528	A/G	0.58	0.50	0.39	0.49	0.46	0.50	0.42	0.43	0.42	0.41	0.40	0.36	0.39	0.38
54	rs1528460	C/T	0.52	0.50	0.80	0.50	0.70	0.50	0.47	0.47	0.42	0.48	0.36	0.38	0.39	0.43

Heterozygosity values were found to be similar across all populations at several loci but there were marked differences seen in some populations especially the Oceanians where several loci showed very low heterozygosity values (e.g. SNP marker code 2, 3 and few others), most likely due to small sample numbers.

Overall, heterozygosity values for the Malays and Chinese were similar to the East Asians, except at few highlighted values (in red), whereas heterozygosity values for the Indians were quite similar to the Central-South Asians. But, there are big differences between both at the highlighted locus.

Table 5-6, demonstrates average observed heterozygosities obtained from Malaysian ethnic groups which were compared with data obtained from (Sanchez et al. 2006a). The results show that the Malaysian Indian ethnic group has similar H_e to the Europeans, and somewhat higher than the Asians and Somali, whilst, the Malays and Chinese show higher He values than the Asians but very close to the Somalis.

Table 5-6: Average Expected heterozygosity (H_e) values of all 52 SNPs markers in different populations, p= 0.05.

	Malaysian[a]			**Asians**[b]	**Somali**[b]	**European**[b]
Average He	***Mal***	***Ch***	***Ind***			
	0.41	0.40	0.44	0.38	0.41	0.44

a: represents average He calculated from this study. Mal= Malays, Ch= Chinese, Ind= Indian
b:represents average He from reference data (Sanchez et al. 2006a).

5.3.3. Exact test for Linkage Equilibrium (LE)

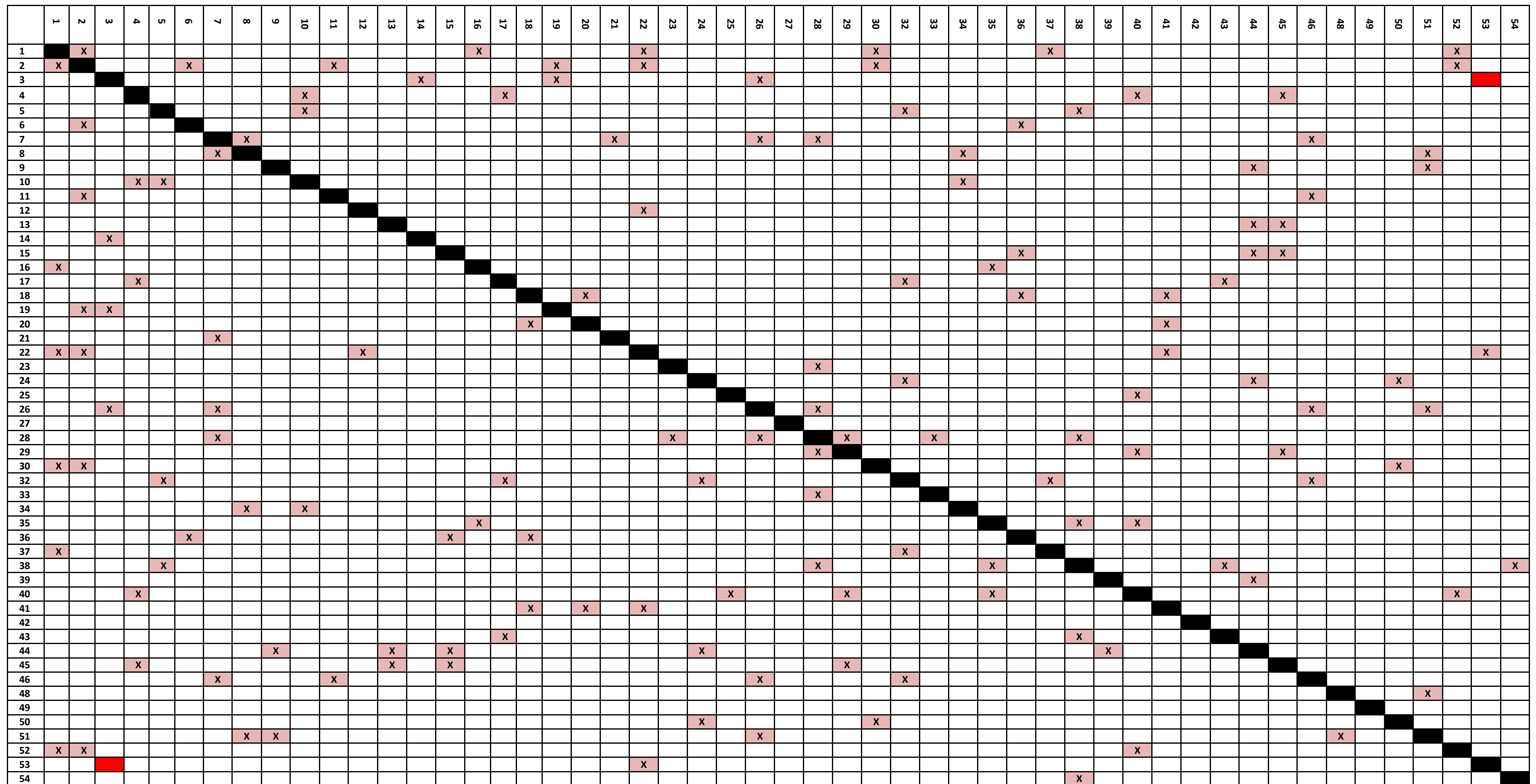
Exact tests for LE using pairwise comparisons in each ethnic group for all 52 loci have also been conducted as shown in Table 5-7. Significant LD was observed at p<0.05: 73 pairs of loci in the Malays, 55 pairs in the Chinese and 80 pairs in the Indians.

The significant LD of all pairs of loci was plotted as shown in Figure 5-3 for the Malays, Figure 5-4 for the Chinese and Figure 5-5 for the Indians. All significant LD at p<0.05 are highlighted with the pink boxes.

However, after Bonferroni correction was applied at p<0.0000377 (0.05 divided by 2601 (51x51 loci number of loci involved)) only 1 pair of loci (highlighted with a red box) in the Malays was found to be remain significant (at p<0.00001922). The association is between locus 3 (at Chromosome 3) and locus 53 (at Chromosome 22).

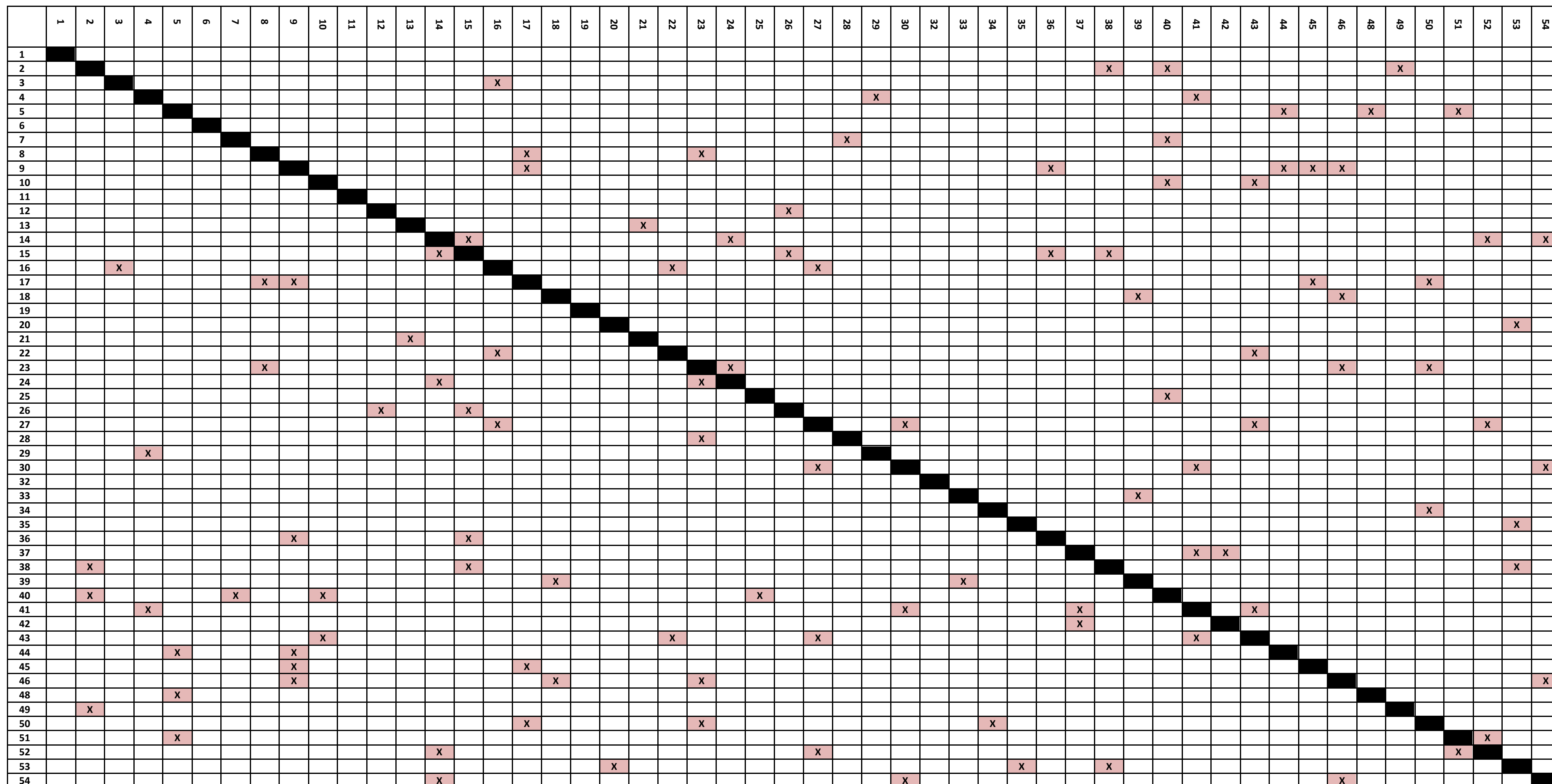
Table 5-7: Significant linkage disequilibrium at $p<0.05$ in Malays, Chinese and Indians.

Marker Code	SNP (rs#)	Linked loci per locus		
		Malay	Chinese	Indian
1	rs1490413	6	0	0
2	rs876724	7	3	2
3	rs1357617	4	1	2
4	rs2046361	4	2	6
5	rs717302	3	3	2
6	rs1029047	2	0	2
7	rs917118	5	1	4
8	rs763869	3	2	3
9	rs1015250	2	5	5
10	rs735155	3	2	3
11	rs901398	2	0	2
12	rs2107612	1	1	1
13	rs1886510	2	1	4
14	rs1454361	1	4	2
15	rs2016276	3	4	4
16	rs729172	2	3	2
17	rs740910	3	4	4
18	rs1493232	3	2	3
19	rs719366	2	0	6
20	rs1031825	2	1	0
21	rs722098	1	1	2
22	rs733164	5	2	6
23	rs826472	1	5	1
24	rs2831700	3	2	2
25	rs873196	1	1	2
26	rs1382387	5	2	3
27	rs2111980	0	4	4
28	rs2056277	6	1	2
29	rs1024116	3	1	3
30	rs727811	3	3	2
32	rs1413212	5	0	2
33	rs938283	1	1	4
34	rs1979255	2	1	4
35	rs1463729	3	1	3
36	rs2076848	3	2	8
37	rs1355366	2	2	6
38	rs907100	5	3	2
39	rs354439	1	2	0
40	rs2040411	5	4	0
41	rs737681	3	4	7
42	rs2830795	0	1	0
43	rs251934	2	4	2
44	rs914165	5	2	3
45	rs10495407	4	2	7
46	rs1360288	4	4	3
48	rs964681	1	1	2
49	rs1005533	0	1	7
50	rs8037429	2	3	3
51	rs891700	4	2	3
52	rs1335873	3	3	2
53	rs1028528	2	3	2
54	rs1528460	1	3	6
Total linked loci		**146 (73 pairs)**	**110 (55 pairs)**	**160 (80 pairs)**



Note: [x] represent significant LD with p<0.05, but not significant after Bonferroni correction at p<0.00001922.
[red box] represent significant LD d after Bonferroni correction at p<0.00001922.

Figure 5-3: Significant LD detected in the Malays.



Note: x represent significant LD with p<0.05, but not significant after Bonferroni correction at p<0.00001922.

Figure 5-4: Significant LD detected in the Chinese.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	48	49	50	51	52	53	54
1 |■|																																																			
2		■							X																						X																							
3			■	X									X																																								
4			X	■				X							X											X									X	X																	
5					■													X																					X																
6						■																X																														X	
7							■																	X							X								X				X							
8				X				■																																									X				X
9		X							■										X			X					X										X															
10										■								X																					X						X							
11											■										X												X																				
12												■																																								X
13			X											■																																	X			X				X
14														■					X								X																										
15				X											■	X								X																												X		
16																X	■																																X					
17																		■																					X				X				X			X					
18					X					X								■	X																																		
19									X					X				X	■									X				X																					
20																					■																																	
21											X										■																														X
22						X			X													■										X														X					X	X	
23																							■																							X				
24															X										■										X															
25							X																			■																				X					
26				X										X													■								X															
27									X																			■									X			X									X			
28																			X									■														X								
29		X																											■						X							X									
30																														■			X												X						
32																			X													■															X				
33							X				X										X										■								X											
34																														X			■							X			X				X						
35				X																																■	X	X																
36				X																					X		X			X						X	■										X	X				X			
37									X	X							X																		X		■						X		X									
38					X																						X									■															
39																																					■															
40																																						■														
41							X										X												X					X	X							■							X				X
42																																									■											
43																																					X						■	X										
44																													X					X										X	■									
45							X			X			X				X																					X	X									■								X
46																						X						X								X											■				X			
48																														X																■		X				
49								X					X			X	X									X								X							X							■					
50																								X							X															X		■					
51																			X							X									X														■			
52															X								X																											■		
53						X																X																												■	
54								X				X	X								X																				X				X									■

Note: X represent significant LD with p<0.05, but not significant after Bonferroni correction at p<0.00001922.

Figure 5-5: Significant LD detected in the Indians.

5.3.4. Population Genetic Structure

To examine population differentiation among the Malaysian ethnic groups, pairwise F_{ST} values were determined using Arlequin version 3.5.1.3, with a total of 1023 permutations for associated p values as shown in Table 5-8. The F_{ST} value between Malay and Chinese at 0.00711 is significantly lower than that between Malay and Indian or between Chinese and Indian (0.0346 and 0.04133 respectively) indicating a closer relationship between Malay and Chinese ethnic groups.

Table 5-8: Pairwise F-statistic (F_{ST}) values of 52 SNPs.

	Malays	Chinese	Indians
Malays	0.00000		
Chinese	0.00711	0.00000	
Indians	0.03460	0.04133	0.00000

F-statistic (F_{ST}) p values.

	Malays	Chinese	Indians
Malays	*		
Chinese	0.00000 +- 0.0000	*	
Indians	0.00000 +- 0.0000	0.00000 +- 0.0000	*

5.3.5. Population assignment of individual genotypes

The likelihood of individual multi-locus genotypes belonging to each of the three groups was analyzed using both Snipper and Arlequin ver 3.5 software. The results are presented in Table 5-9. Results from both software (A and B) gave very similar proportions of distribution of the individuals within each ethnic group. Within Malays, >

70% of the genotypes was correctly classified whereas >81 % and >88 % were correctly classified within Chinese and Indian ethnic groups. A larger proportion of Malay genotypes were misclassified as Chinese (17 – 21%) than Indian (8%). Similarly, a larger proportion of Chinese were misclassified as Malay (18 – 19%) than Indian (1- 3%). The proportion of misclassified genotypes was minimal in the Indian group (11%) but interestingly, a greater proportion was misclassified as Malay (8%) than Chinese (2-3%).

Table 5-9: Percentages of population assignment using Snipper (A) and Arlequin version 3.5 (B).

Likelihood	Malays		Chinese		Indians	
	A	B	A	B	A	B
Population of Malay origin	72.48%	70.64%	19.27%	17.76%	8.26%	8.26%
Population of Chinese origin	16.82%	21.10%	81.31%	81.31%	1.87%	2.75%
Population of Indian origin	8.26%	8.26%	2.75%	0.93%	88.99%	88.99%

The classification probabilities obtained are also represented as a triangular plot and density plots in Figure 5-6, which (A) clearly shows many Malay and Chinese individuals positioned closely together. In contrast, fewer Chinese and Malay were misclassified as Indians, and Indians as Chinese and Malay. Once again, this indicates a close relationship between the Malays and the Chinese, than either has to the Indians. Whilst, Figure 5-6 (C), (D) and (E) show each group density plot and (B) is when the group plots were overlaid on top of each other.

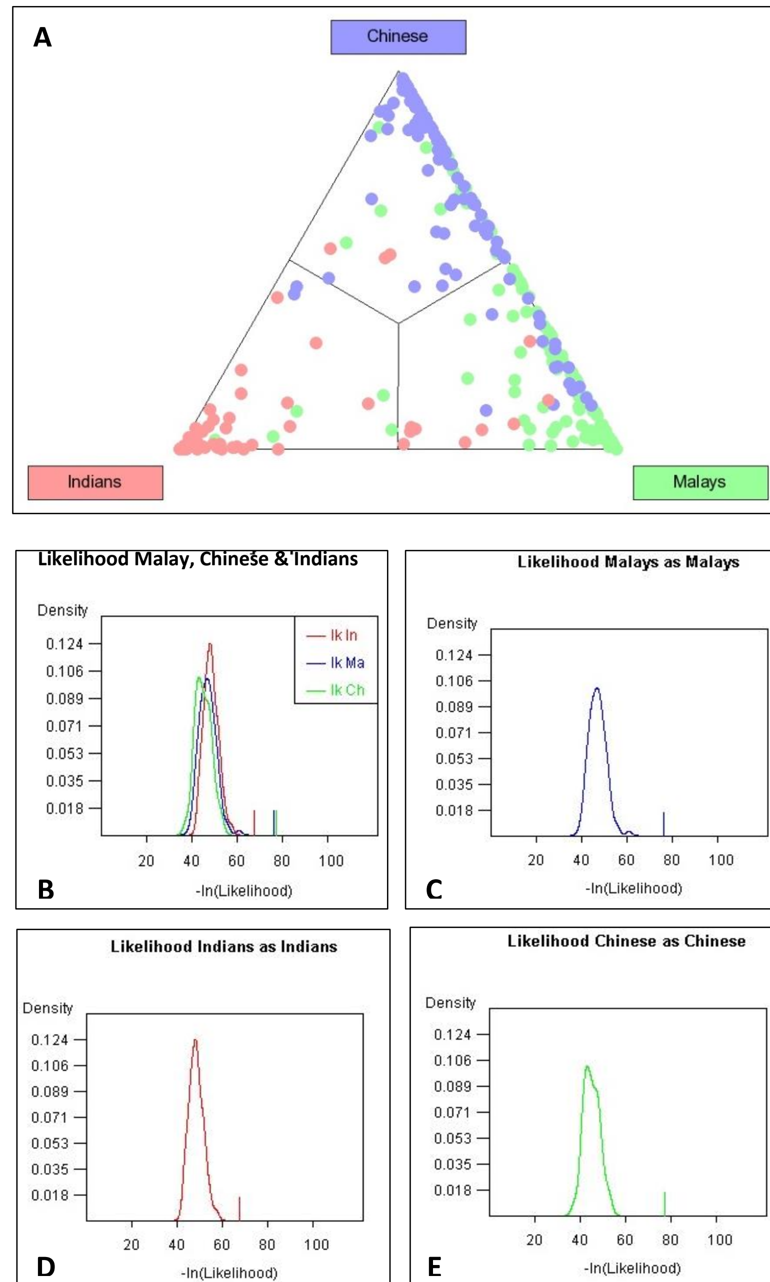


Figure 5-6: (A) Triangular plot classifying the individuals, (B) to (E) are density plot of the likelihood of all ethnic groups of data analyzed with the Snipper software.

Table 5-10: Use of 'cross verbose-validation' to compute the success ratio using the 52 SNPs with The Snipper software.

Success ratio computation			
	Malays	**Chinese**	**Indians**
Population of Malay origin	**22.94%**	60.55%	16.51%
Population of Chinese origin	0.00%	**53.27%**	46.73%
Population of Indian origin	5.50%	0.00%	**94.50%**

Table 5-10 shows the performance of the 52 SNPs that were used in this study in differentiating the co-ancestry contributors. The results show that the success ratio (values in bold) for the Indians is far higher at 94.5% than in the Chinese (53.27%) and the Malays (22.94%). Indirectly, this also indicates these 52 SNP markers have high probabilities in discriminating Indians from the Malays and the Chinese with only a low probability of differentiating between the Chinese and Malays.

5.3.6. Forensic statistical analysis

The match probability (MP) was calculated for the 52 locus profile using allele frequency data obtained from each ethnic group as shown in Table 5-11.

The combined mean MP for the 52 SNPs markers of Malays, Chinese, and Indians are 1 in $3.9654e^{-19}$, 1 in $5.3964e^{-18}$ and 1 in $1.7459e^{-19}$ respectively corresponding to a combined power of discrimination of >99.9999% in each case..

Most of the SNP markers showed PIC values (Botstein et al. 1980) greater than 0.25, which shows that these SNP markers are reasonably informative for Malaysian population samples.

In Malays, 9 markers were found to be less informative: code marker 3, 13, 16, 17, 25, 28, 29, 41 and 43, in the Chinese, 12 markers were less informative: code marker 5, 10, 16, 17, 22, 23, 25, 28, 29, 33, 41 and 43, whilst in the Indians, only four markers were appeared less informative: code marker 16, 17, 33 and 45. These are highlighted in red, emboldened and italicized.

It is also notable that the least informative marker is SNP marker 29 in Malay and Chinese groups. All three ethnic groups show low informativeness at 2 SNP markers: code marker 16 and 17.

Table 5-12 demonstrates typical paternity index (PI) and power of exclusion (PE) for each of the 52 loci within each ethnic group. The combined PI (CPI) and average exclusion probability were also estimated. Results showed that the CPI for Malay, Chinese and Indians are $9.7072e^{-4}$, $1.9363e^{-2}$ and $8.7491e^{-2}$, respectively. The exclusion probabilities for all ethnic groups are similar at 99.99%.

Table 5-11: Forensic parameters calculated for 52 SNPs from each Malaysian ethnic group. MP represents match probability, PD represents power of discrimination, and PIC represents polymorphism information content.

Marker Code	SNP (rs#)	Allele (1/2)	Malay			Chinese			Indian		
			MP	PD	PIC	MP	PD	PIC	MP	PD	PIC
1	rs1490413	T/C	0.379	0.621	0.37	0.351	0.649	0.37	0.383	0.617	0.37
2	rs876724	C/T	0.420	0.580	0.35	0.409	0.591	0.34	0.404	0.596	0.34
3	rs1357617	T/A	0.567	0.433	**0.23**	0.543	0.457	0.25	0.474	0.526	0.32
4	rs2046361	T/A	0.373	0.627	0.37	0.404	0.596	0.37	0.377	0.623	0.36
5	rs717302	A/G	0.433	0.557	0.31	0.607	0.393	***0.22***	0.419	0.581	0.36
6	rs1029047	A/T	0.397	0.603	0.35	0.431	0.569	0.33	0.357	0.643	0.37
7	rs917118	G/A	0.485	0.515	0.29	0.457	0.543	0.32	0.419	0.581	0.37
8	rs763869	C/T	0.355	0.645	0.37	0.373	0.627	0.37	0.501	0.499	0.28
9	rs1015250	G/C	0.371	0.629	0.37	0.364	0.636	0.37	0.369	0.631	0.37
10	rs735155	T/C	0.546	0.454	0.25	0.628	0.372	***0.20***	0.397	0.603	0.35
11	rs901398	C/T	0.456	0.544	0.31	0.418	0.582	0.33	0.402	0.598	0.35
12	rs2107612	A/G	0.442	0.558	0.33	0.467	0.533	0.31	0.563	0.437	0.38
13	rs1886510	G/A	0.593	0.407	***0.21***	0.566	0.434	***0.23***	0.453	0.547	0.32
14	rs1454361	A/T	0.369	0.631	0.37	0.409	0.591	0.37	0.383	0.617	0.36
15	rs2016276	T/C	0.391	0.609	0.35	0.395	0.605	0.35	0.447	0.553	0.31
16	rs729172	T/G	0.579	0.421	**0.23**	0.610	0.390	***0.21***	0.627	0.373	***0.21***
17	rs740910	A/G	0.609	0.391	**0.20**	0.720	0.280	***0.14***	0.575	0.425	***0.23***
18	rs1493232	T/G	0.408	0.592	0.36	0.417	0.583	0.36	0.378	0.622	0.36
19	rs719366	C/T	0.423	0.577	0.35	0.496	0.504	0.28	0.416	0.584	0.34
20	rs1031825	T/G	0.390	0.610	0.37	0.399	0.601	0.37	0.414	0.586	0.37
21	rs722098	A/G	0.391	0.609	0.37	0.399	0.601	0.37	0.614	0.386	0.37
22	rs733164	A/G	0.436	0.564	0.32	0.575	0.425	***0.24***	0.534	0.466	0.25
23	rs826472	G/A	0.511	0.489	0.27	0.638	0.362	***0.19***	0.378	0.622	0.36
24	rs2831700	C/T	0.379	0.621	0.37	0.402	0.598	0.37	0.486	0.514	0.29
25	rs873196	A/G	0.580	0.420	***0.23***	0.717	0.283	***0.15***	0.528	0.472	0.26
26	rs1382387	G/T	0.528	0.472	0.37	0.447	0.553	0.36	0.440	0.560	0.37
27	rs2111980	A/G	0.411	0.589	0.36	0.378	0.622	0.36	0.398	0.602	0.37
28	rs2056277	A/G	0.522	0.448	***0.24***	0.652	0.348	***0.19***	0.506	0.494	0.27
29	rs1024116	A/G	0.706	0.294	***0.16***	0.733	0.267	***0.14***	0.426	0.574	0.32
30	rs727811	A/C	0.386	0.614	0.37	0.403	0.597	0.35	0.459	0.541	0.34
32	rs1413212	C/T	0.457	0.543	0.37	0.415	0.585	0.35	0.476	0.524	0.36
33	rs938283	C/T	0.434	0.566	0.32	0.597	0.403	***0.22***	0.575	0.425	***0.23***
34	rs1979255	G/C	0.354	0.646	0.37	0.370	0.630	0.37	0.381	0.619	0.36
35	rs1463729	A/G	0.382	0.618	0.37	0.370	0.630	0.37	0.414	0.586	0.37
36	rs2076848	A/T	0.470	0.530	0.31	0.420	0.580	0.34	0.550	0.450	0.37
37	rs1355366	C/T	0.440	0.560	0.34	0.560	0.440	***0.24***	0.388	0.612	0.37
38	rs907100	G/C	0.336	0.664	0.37	0.349	0.651	0.37	0.343	0.647	0.37
39	rs354439	A/T	0.365	0.635	0.37	0.400	0.600	0.37	0.366	0.634	0.37
40	rs2040411	A/G	0.414	0.586	0.34	0.401	0.599	0.35	0.412	0.588	0.37
41	rs737681	C/T	0.560	0.440	***0.24***	0.609	0.391	***0.21***	0.367	0.633	0.36
42	rs2830795	A/G	0.398	0.602	0.36	0.434	0.566	0.37	0.436	0.564	0.33
43	rs251934	A/G	0.625	0.375	***0.20***	0.666	0.334	***0.18***	0.428	0.572	0.34
44	rs914165	C/T	0.419	0.581	0.33	0.384	0.616	0.35	0.369	0.631	0.37
45	rs10495407	C/T	0.433	0.567	0.32	0.422	0.578	0.33	0.571	0.429	***0.23***
46	rs1360288	A/G	0.490	0.510	0.28	0.524	0.476	0.35	0.417	0.583	0.37
48	rs964681	C/T	0.408	0.592	0.35	0.432	0.568	0.35	0.377	0.623	0.37
49	rs1005533	A/G	0.517	0.483	0.27	0.442	0.558	0.33	0.432	0.568	0.35
50	rs8037429	A/G	0.394	0.606	0.37	0.443	0.557	0.36	0.372	0.628	0.36
51	rs891700	A/G	0.379	0.621	0.37	0.351	0.649	0.37	0.416	0.584	0.37
52	rs1335873	T/A	0.443	0.557	0.32	0.405	0.595	0.36	0.418	0.582	0.33
53	rs1028528	A/G	0.429	0.571	0.37	0.354	0.646	0.37	0.362	0.638	0.37
54	rs1528460	A/G	0.389	0.611	0.37	0.666	0.334	0.37	0.553	0.467	0.37
Combined calculations			**MP = 1 in 3.9654e^{-19} PD >99.9999%**			**MP = 1 in 5.3964e^{-18} PD >99.9999%**			**MP = 1 in 1.7459e^{-19} PD >99.9999%**		

Table 5-12: Combined Paternity index and exclusion probability of 52 markers in Malays, Chinese and Indian groups.

Marker Code	SNP (rs#)	Allele (1/2)	Malay		Chinese		Indian	
			PI	PE	PI	PE	PI	PE
1	rs1490413	T/C	0.97	0.176	0.86	0.127	1.03	0.200
2	rs876724	C/T	1.01	0.192	0.85	0.121	0.85	0.122
3	rs1357617	T/A	0.70	0.057	0.71	0.063	0.99	0.184
4	rs2046361	T/A	0.96	0.176	1.09	0.227	0.89	0.140
5	rs717302	A/G	0.75	0.077	0.64	0.034	1.09	0.226
6	rs1029047	A/T	0.87	0.128	0.85	0.121	0.92	0.154
7	rs917118	G/A	0.74	0.073	0.96	0.168	1.11	0.236
8	rs763869	C/T	0.87	0.128	0.94	0.160	0.83	0.111
9	rs1015250	G/C	0.94	0.161	0.96	0.168	0.89	0.140
10	rs735155	T/C	0.68	0.050	0.65	0.040	0.87	0.128
11	rs901398	C/T	0.81	0.105	0.76	0.084	0.88	0.134
12	rs2107612	A/G	0.92	0.154	0.91	0.146	1.82	0.468
13	rs1886510	G/A	0.70	0.057	0.73	0.071	0.88	0.134
14	rs1454361	A/T	0.94	0.161	1.09	0.227	0.92	0.154
15	rs2016276	T/C	0.87	0.128	0.89	0.140	0.74	0.073
16	rs729172	T/G	0.67	0.047	0.64	0.037	0.61	0.023
17	rs740910	A/G	0.68	0.050	0.60	0.022	0.66	0.044
18	rs1493232	T/G	1.05	0.208	1.03	0.200	0.91	0.147
19	rs719366	C/T	0.97	0.176	0.75	0.080	0.87	0.128
20	rs1031825	T/G	1.05	0.208	1.09	0.227	1.14	0.245
21	rs722098	A/G	1.01	0.192	1.09	0.227	2.10	0.530
22	rs733164	A/G	0.83	0.111	0.65	0.040	0.73	0.069
23	rs826472	G/A	0.77	0.086	0.64	0.037	0.91	0.147
24	rs2831700	C/T	0.97	0.176	1.09	0.227	0.78	0.090
25	rs873196	A/G	0.65	0.038	0.59	0.020	0.74	0.073
26	rs1382387	G/T	1.56	0.396	1.19	0.267	1.18	0.265
27	rs2111980	A/G	1.05	0.208	0.92	0.153	1.09	0.266
28	rs2056277	A/G	0.74	0.073	0.61	0.026	0.76	0.081
29	rs1024116	A/G	0.59	0.019	0.59	0.020	0.75	0.077
30	rs727811	A/C	1.03	0.200	0.96	0.168	1.09	0.226
32	rs1413212	C/T	1.27	0.298	0.96	0.168	1.30	0.309
33	rs938283	C/T	0.84	0.116	0.64	0.037	0.66	0.044
34	rs1979255	G/C	0.89	0.140	0.96	0.168	0.88	0.134
35	rs1463729	A/G	1.03	0.200	0.97	0.176	1.14	0.245
36	rs2076848	A/T	0.88	0.134	0.92	0.153	1.70	0.438
37	rs1355366	C/T	1.03	0.200	0.74	0.075	1.05	0.208
38	rs907100	G/C	0.77	0.086	0.88	0.133	0.68	0.050
39	rs354439	A/T	0.88	0.134	1.05	0.209	0.92	0.154
40	rs2040411	A/G	0.91	0.147	0.86	0.127	1.14	0.245
41	rs737681	C/T	0.71	0.061	0.67	0.046	0.87	0.128
42	rs2830795	A/G	1.01	0.192	1.22	0.278	0.89	0.140
43	rs251934	A/G	0.63	0.033	0.61	0.026	0.92	0.154
44	rs914165	C/T	0.77	0.086	0.86	0.127	0.92	0.154
45	rs10495407	C/T	0.80	0.100	0.86	0.127	0.73	0.069
46	rs1360288	A/G	0.75	0.077	1.41	0.348	1.16	0.255
48	rs964681	C/T	0.91	0.147	1.07	0.218	0.97	0.176
49	rs1005533	A/G	0.74	0.073	0.91	0.146	1.01	0.192
50	rs8037429	A/G	1.03	0.200	1.14	0.246	0.88	0.134
51	rs891700	A/G	1.01	0.192	0.88	0.133	1.11	0.236
52	rs1335873	T/A	0.89	0.140	1.01	0.192	0.80	0.100
53	rs1028528	A/G	1.18	0.265	0.82	0.109	0.92	0.154
54	rs1528460	A/G	1.05	0.208	2.55	0.606	1.65	0.424
Cumulative probability			**CPI= 9.7072e^{-4} Ave. PE= 99.99%**		**CPI= 1.9363e^{-2} Ave. PE=99.99%**		**CPI= 8.7491e^{-2} Ave. PE=99.99%**	

All forensic data obtained from the Malays, Chinese and Indians were also compared to other reference population data (SNPforID Consortium) as shown in Table 5-13.

The Europeans had the lowest cumulative MP values across 52 loci compared to the others. The Malaysians (Malays, Chinese and Indians) had similar RMP with the East Asians, East Timorese and the Africans. The Oceanians showed slightly high matching probability value from the others.

However, all populations show identical combined power of discrimination, which was 99.9999 % for all loci, and the Malaysian ethnic groups demonstrate the highest percentage of cumulative exclusion probability.

Table 5-13: Forensic data comparison among the populations.

Parameters	**Populations**							
	Malaysia			**East Asia[1]**	**East Timor[1]**	**Oceania[1]**	**Africa[1]**	**Europe[1]**
	Malay	**Chinese**	**Indian**					
Cumulative random match probability	**4.0×10^{-19}**	**5.4×10^{-18}**	**1.7×10^{-19}**	**5.0×10^{-19}**	**2.0×10^{-18}**	**5.2×10^{-16}**	**1.1×10^{-19}**	**5.0×10^{-21}**
Combined power of discrimination	**99.9999%**	**99.9999%**	**99.9999%**	**99.9999%**	**99.9999%**	**99.9999%**	**99.9999%**	**99.9999%**
Cumulative exclusion probability	**99.99%**	**99.99%**	**99.99%**	**99.91%**	**99.87%**	**99.58%**	**99.95%**	**99.98%**

1: data obtained from (Santos et al. 2011, SNPforID Consortium)

5.3.7. Comparison of forensic parameters between Malaysian SNP data with Malaysian STR data

SNP data obtained in this study were compared to the existing Malaysia STR allele frequency database at Department of Chemistry Malaysia. The results show that the 52 SNPs data gave a lower matching probability compared to the STRs. However, the rest of the parameters remain similar (shown in Table 5-14).

Table 5-14: Forensic data collected and estimated for Malaysian populations.

Population	Cumulative random match probability	Combined power of discrimination	Cumulative exclusion probability
52-SNPs:			
Malay	4.10×10^{-19}	99.9999%	99.99%
Chinese	5.4×10^{-18}	99.9999%	99.99%
Indians	1.7×10^{-19}	99.9999%	99.99%
STR-Peninsular Malaysia[a]			
Malay	2.6×10^{-17}	99.9999%	99.99%
Chinese	7.0×10^{-16}	99.9999%	99.99%
Indians	3.6×10^{-16}	99.9999%	99.99%
STR-East Malaysia (Sarawak)[b]			
Iban	3.29×10^{-17}	99.9999%	99.99%
Bidayuh	6.52×10^{-17}	99.9999%	99.99%
Melanau	1.52×10^{-16}	99.9999%	99.99%

a: data obtained from (Seah et al. 2003) , and
b: data obtained from (Suadi et al. 2007).

5.5. DISCUSSION

The increasing interest in using SNPs as forensic markers has led to more population studies. However, allele frequencies can vary greatly among populations; therefore it is essential to determine the allele distributions of relevant SNPs within populations before forensic analysis. These SNPs can also provide valuable information for population admixture detection and in the estimation of bio-geographical ancestry (Sanchez et al. 2006a).

5.5.1. Allele distribution among the populations

Although the 52 SNP markers used in this study were chosen by Sanchez et al (2006) because they had allele frequencies close to 0.5/0.5 in Europeans, frequencies observed in Malay, Chinese and Indian ethnic groups varied, ranging from a minimum of 0.079 and a maximum of 0.921.

The Malays and the Chinese showed very similar allele frequencies for most SNP loci used in this study, and both showed similar frequencies to the East Asians (SNPforID Consortium). However, the Malaysian Indians appeared to be more similar with the Central-South Asians.

5.5.2. Hardy-Weinberg Equilibrium (HWE) and LE departures

Among the Malaysian groups, the Chinese demonstrated the lowest average heterozygosity across the 52 loci (0.40). Sanchez also has reported that among the population groups studied, the lowest heterozygosity was found in the Greenlander and Taiwanese groups, which both with 0.37. In contrast, the Malaysian Indians had the highest heterozygosity (0.46) among the populations even higher than the Europeans (0.44). In the original paper, Sanchez and colleagues have stated that they expected to obtained the highest heterozygosity value for the Europeans because one

of the primary SNP selection criteria was for maximum heterozygosity in European population with less emphasis on other populations.

As for pairwise LE analysis, after the Bonferroni correction ($p<0.00001922$), there was only one significant association between loci was found between the SNP marker 3 and SNP marker 53 in the Malays. However, two other groups have no significant deviation. In the original study (Sanchez et al. 2006a), no significant association of the alleles across 52 loci was reported after sequential Bonferroni corrections ($p>0.05$).

5.5.3. Pairwise FST estimates between the populations

Apart from having different HWE and LE departure at different SNP markers for all three ethnic groups, the Malays, the Chinese and the Indians, the F_{ST} evaluations also show significant differences between these three groups. The Malays were found to be more closely related to the Chinese (with a F_{ST} of 0.007) than either was to the Indians (F_{ST} = 0.034 and 0.041 respectively).

5.5.4. The relationship between the Malays, the Chinese and the Indians

Results obtained from this study indicate that the Malays are more closely related to the Chinese, compared to the Malays with the Indians or the Chinese with the Indians. The population assignment results also supported the above findings (table 5-9 and Figure 5-6). The HUGO Pan-Asian SNP Consortium having studied thousands of autosomal SNPs, suggested that the East Asians and Southeast Asians share a common origin, whilst the Central-South Asians share ancestry with the Europeans (Abdulla et al. 2009). In addition, Hatin and colleagues have also revealed their findings after analyzing approximately 50, 000 autosomal SNPs and described the close connections between the Malays, the Chinese, the Indonesian and the Indigenous people in the phylogenetic tree (Hatin et al. 2011).

Furthermore, Chang and colleagues have carried out a population study using 16 Y-STR markers on 980 male individuals from these three major ethnic groups and the results showed that the Malay and the Chinese populations were more closely related to each other than either was to the Indians (Chang et al. 2007). These findings were supported by Yong et al. (2006), who reported both groups, the Malays and the Chinese in Singapore as sharing most common Y-STR alleles than they did with the Indians (Yong et al. 2006). Even the frequency data reported in both countries correlates with those published for the Han Chinese (Hou et al. 2001) and the South Indian populations (Ballard et al. 2005).

However, another interesting finding by Chang and colleagues (2007) that support the SNP population assignment in this study (Table 5-9 and Figure 5-6) was the Malays appeared to be more closely related to the Indians than the Chinese, when the samples were analyzed with the additional loci that were included in the Applied Biosystems Yfiler® kit. These additional loci in the Y-filer® system: DYS448, DYS635 and DYS458 are the three most diverse single loci for the three ethnic populations (Chang et al. 2007). This finding shows that although Indians are quite different, they are somewhat closer to Malays than Chinese.

CHAPTER 6

ANALYSIS OF 13-PLEX SNAPSHOT ASSAYS ON THE FORENSIC BIOLOGICAL SAMPLES

PART ONE:

KINSIHP ANALYSIS

PART TWO:

(A) SIMULATED CRIME SAMPLES

(B) REAL CASEWORK SAMPLES

ANALYSIS OF 13-PLEX SNAPSHOT ASSAYS ON THE FORENSIC BIOLOGICAL SAMPLES

6.1. INTRODUCTION

There are many challenges involved with the amplification of crime-scene samples for forensic STR profiling. Samples may contain inhibitory substances, be heavily degraded, have only trace quantities of biological material, or be from multiple donors. Often minute quantities of DNA template will result partially amplified STR profiles. Partial profiles, amplification bias and allele drop-in can all be observed with compromised samples. While current multiplexes perform exceptionally well with a wide range of samples, new methods and technologies are being developed which could further increase the success rates of current STR profiling reactions from compromised samples (Ballantyne et al. 2011).

The sole application of STR analysis in paternity testing can be problematic in various ways, and might lead to ambiguous results which due a low probability of relationship or when an exclusion of fatherhood is suggested by just 1 or 2 STR loci, e.g. when (unknowingly) testing close relatives of the alleged father. In addition, the STR results provide insufficient information for an exclusion because of the relatively frequently occurring germ-line step mutations in STR loci (addition or the decrease of 1 or 2 repeats) that cannot be distinguished from exclusions (Schwark et al. 2012a).

6.2. RATIONALE OF THE STUDY

In order to evaluate the reliability of the 13-plex assays for the real casework samples, several studies were conducted. The studies included the relationship testing in the paternity cases, and the analysis of the simulated and real crime samples. All data obtained were analyzed with DNA.VIEW^{TM} software.

6.2.1. Section One: Kinship Analysis

For this section, 10 real kinship cases samples and 2 known family samples were used. Seven of them were trios, four duos and a family of five (known control). In two of the cases, there are genetic inconsistencies at one or two loci, which could be due to mutation. All these samples were previously amplified with AmpF1STR Identifiler® kit (Applied Biosysytems) and also were re-amplified with the Powerplex 16® System (Promega) for comparison study.

6.2.2. Section Two: Crime Samples

For this section, the 13-plex assays were subjected to two types of samples: environmental simulated samples and real crime casework samples.

6.2.2.1. Environmental simulated samples

This simulated sample was tested under Malaysian temperate and humidity environment for 27 days continuously as described earlier in *Chapter 2, section 2.1.2*. The samples were collected every day (from Day 1 to Day 27) and were kept in the freezer. Throughout the experiment, the samples were covered with the plastic packet to avoid interferences from any insects or other smaller fauna.

6.2.2.2. Real forensic biological samples

In this study, the 13-plex assays were applied on 51 samples of 10 cases. Most of these samples are bloodstain type, which were obtained from either the crime scene or the victim's belongings. The samples were selected from the casework that has both; failed and successfully amplified with the STR analysis.

6.3. RESULTS

6.3.1. Section One: Kinship Analysis

6.3.1.1. Analysis of different paternity cases (trio, duo and family)

In Case 5, one mutation had occurred at the Child's CSF1PO marker (Powerplex® 16 system) as shown in Figure 6-1. In a control family, the Child E showed two mutations had occurred at two STR loci, D8S1179 and FGA (Figure 6-2). Others are the normal trios and duos paternity/maternity investigations. In all cases, the putative father or mothers are found to be the biological mother or father of the child(s), including in Case 9 (from Table 6-1), AM1 is found to be the biological mother of the child, not AM2.

6.3.1.2. Combined paternity index (CPI) and probability of paternity (%)

The STR analysis results were compared to the SNP genotyping results as shown in Table 6-1.

Overall, in both analyses, the trios showed higher combined paternity index (CPI) and probability of paternity (%) when compared to the duos. In addition, in most of the trio cases involving mother, child and alleged father, the STR gave higher CPI and % than SNP analysis as can be seen in Table 6-1.

However, in cases with one or two mutation occurrences, SNP assays gave higher CPI and % than STR analysis. This demonstrates one of the advantages of SNP technology over the STR analysis.

As for the control that comprised two alleged mothers of a child, the SNP-based probability of maternity gave 100% to AM1 compared to STRs 99.99991%.

Table 6-1: Paternity calculation using 16 STRs versus 52 SNPs markers in all kinship cases.

Case	Remarks	Combined paternity index (CPI)		Probability of paternity (%)	
		16 STRs	52 SNPs	16 STRs	52 SNPs
1	Trio	$1.09e^6$	$2.16e^6$	99.99991	99.99995
2	Duo	179401	796.42	99.9994	99.9
3	Trio	$86.55e^6$	$7.55e^6$	99.999999	99.99999
4	Trio	$43.41e^6$	$9.46e^6$	99.999998	99.99999
5	***Trio- 1 mutation***	***87851***	***$14.15e^6$***	***99.999***	***99.999993***
6	Trio	$11.407e^9$	$328.849e^6$	99.99999999	99.9999997
7	Duo	108.978	$5.14e^6$	99.09	99.99998
8	Control 1-a family of 5 (father and mother) with: Child C Child D ***Child E (2 mutations)***	 $516.24e^6$ $27.72e^6$ ***54.44***	 $3.96e^6$ $325.84e^6$ ***$3.26e^6$***	 99.9999998 99.999996 ***98.2***	 99.99997 99.9999997 ***99.99997***
9	Control 2- 2 AM of a baby girl AM1(M) AM2(G)	 $1.16e^6$ 30.08	 - 1.27	 99.99991 96.8	 100 -
10	Duo	275561	77.32	99.9996	98.7
11	Duo	$5.24e^6$	409.15	99.99998	99.8
12	Trio	$193.90e^6$	$17.43e^6$	99.9999995	99.999994

Notes: AM represents alleged mother.

The DNA.VIEW calculation results are attached as *Appendix F (52 SNPs) and Appendix G (STR)*.

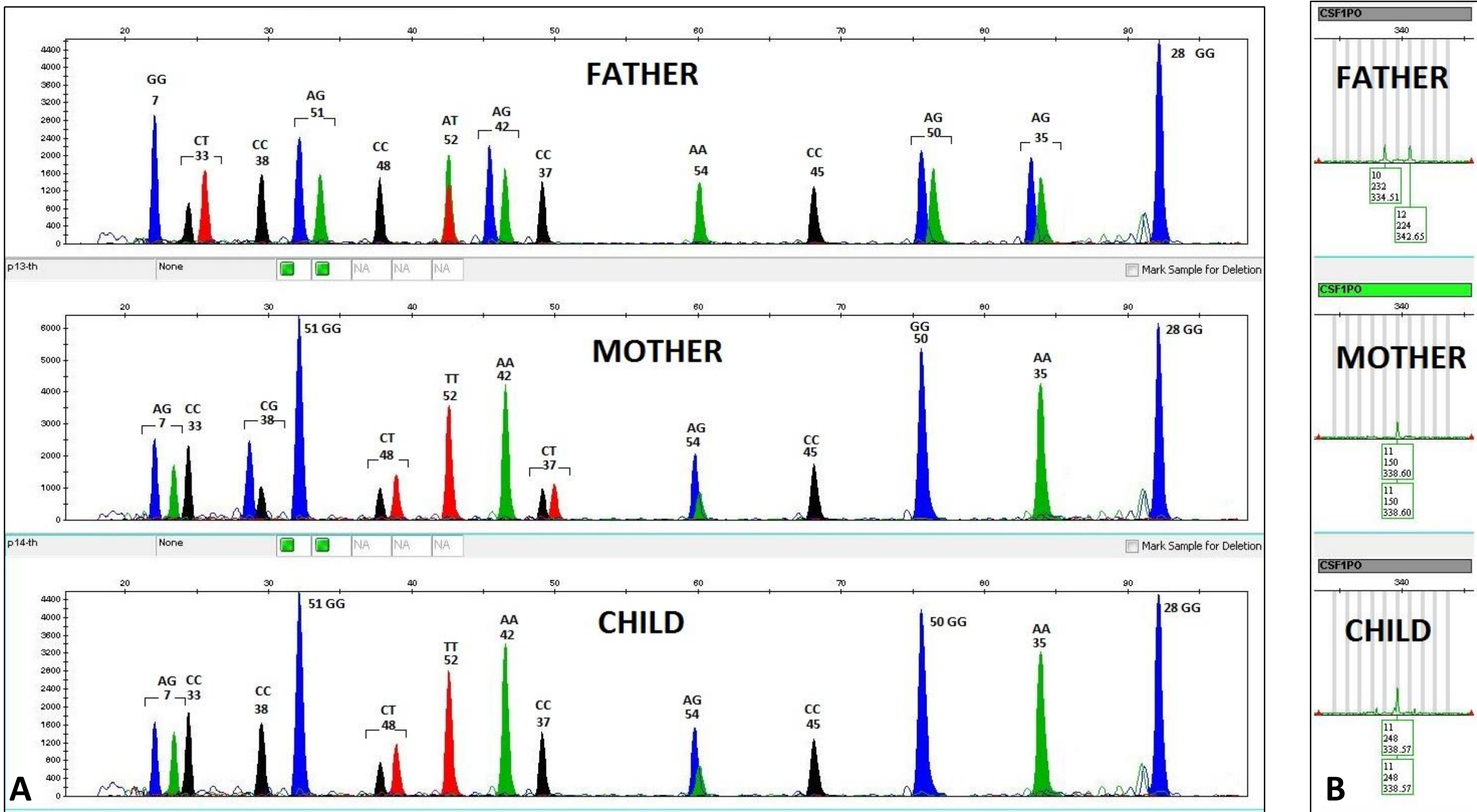


Figure 6-1: Electropherograms of Paternity Case 5 (mentioned in Table 6-1) involving putative father, mother and child: (A) SNaPshot profiles of 13_{th} multiplex assay of this trio, and (B) STR profile with one mutation occurrence at Child's *CSF1PO* marker.

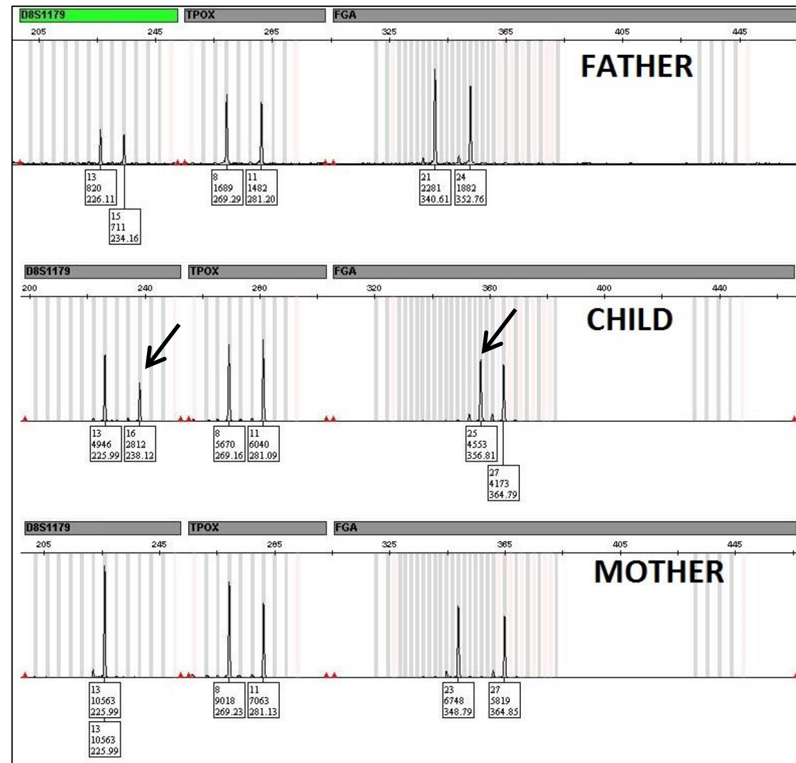


Figure 6-2: STR profile of a Control family (Case 8- Table 6-1) involving father, mother and Child E, where two mutations had occurred in Child E at locus D8S1179 and FGA (as shown by the arrows).

6.3.2. Section Two: Analysis of the crime samples

6.3.2.1. SNPs analysis of the simulated casework samples

Total duration of samples collection was 27 days in the month of August/September. Data of average temperature, rainfall and relative humidity of each day involved (Figure 6-3) have been provided by the Meteorology Department of Malaysia (as shown in *Appendix C*).

As shown in Figure 6-3, the temperature of all 27 days shows an average of 33°C, whilst the humidity varies from day 1 to day 27 and morning (a.m.) to evening (p.m.). Out of 27 days, rainfall was only observed in 11 days with the maximum rainfall 29.4 mm (in Day 19).

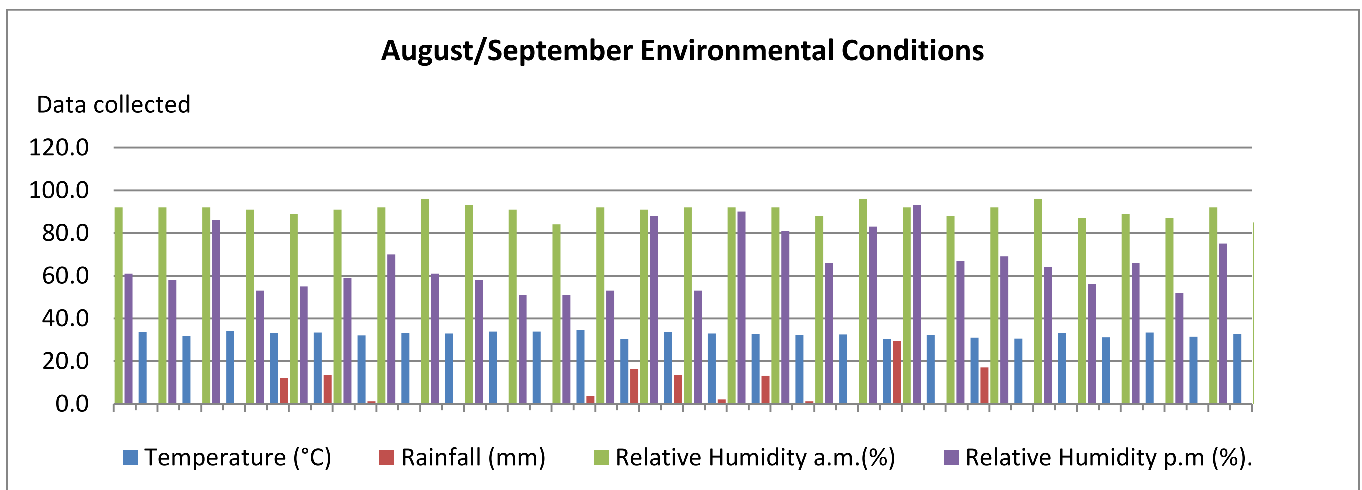


Figure 6-3: Shown above are the parameters such as temperature, rainfall and relative humidity values in the month of August/September (for 27 days).

From Day 22 until Day 27, no rainfall was received (Figure 6-3), however fungi were found surrounded the bloodstain spots on cloth taken from Day 21 to Day 27 as shown in Figure 6-4. This could be because dry temperature gave good conditions for the fungus to grow after the samples were exposed to the temperature, varying humidity and rainfall during the earlier days.

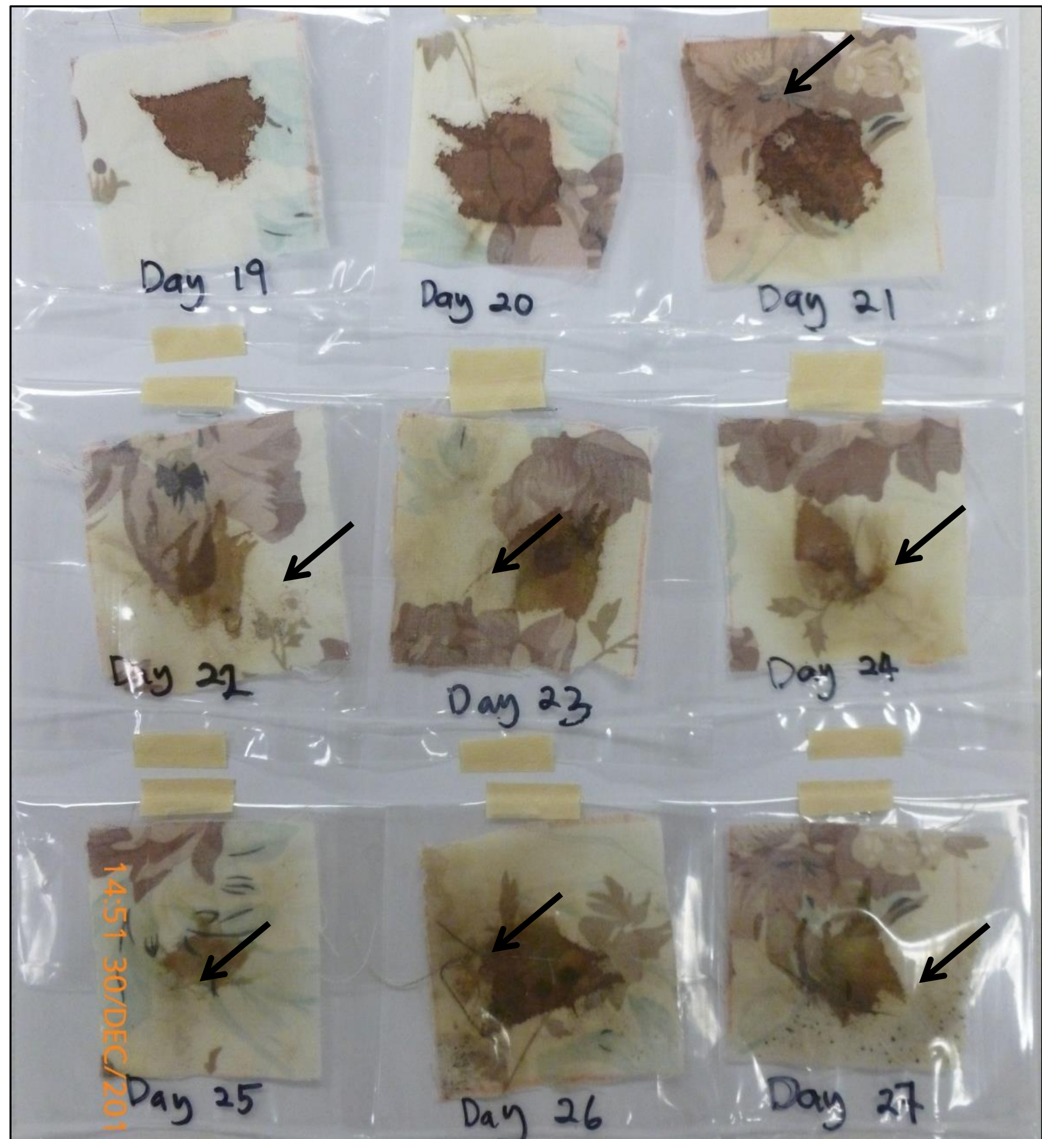


Figure 6-4: Shown above are the simulated samples taken from Day 19 to Day 27. The arrows show the bloodstains on the cloths surrounded or covered by the fungi.

As for quantification results (Table 6-2), DNA was detected until Day 21, and a very minute amount in Day 23, Day 24 and Day 25. Full STR and SNP profiles were obtained from Day 1 to Day 21. Starting from Day 22 to Day 27, only a few full alleles were detected. This might be because of the interference from the fungus.

Table 6-2: Full alleles detected in Day 1, Day 21 to Day 27.

Day samples were taken	**Quantification values (ng/µl)**	**Full alleles detected**	
		Powerplex® 16	52 SNP markers
Day 1	3.45	16	52
Day 21	0.55	16	52
Day 22	Undetermined	0	8
Day 23	0.002	0	11
Day 24	0.003	1	4
Day 25	0.003	0	3
Day 26	Undetermined	5	2
Day 27	Undetermined	0	0

As can be seen from Figure 6-5, only Day 1 and Day 21 have full complete profiles for *13*$_{nd}$ assays (SNPs) and STR analysis. Few alleles were detected for both analyses, but most of the alleles that generated from these samples (such as for Day 26 and Day 27) are not the true/correct allele(s).

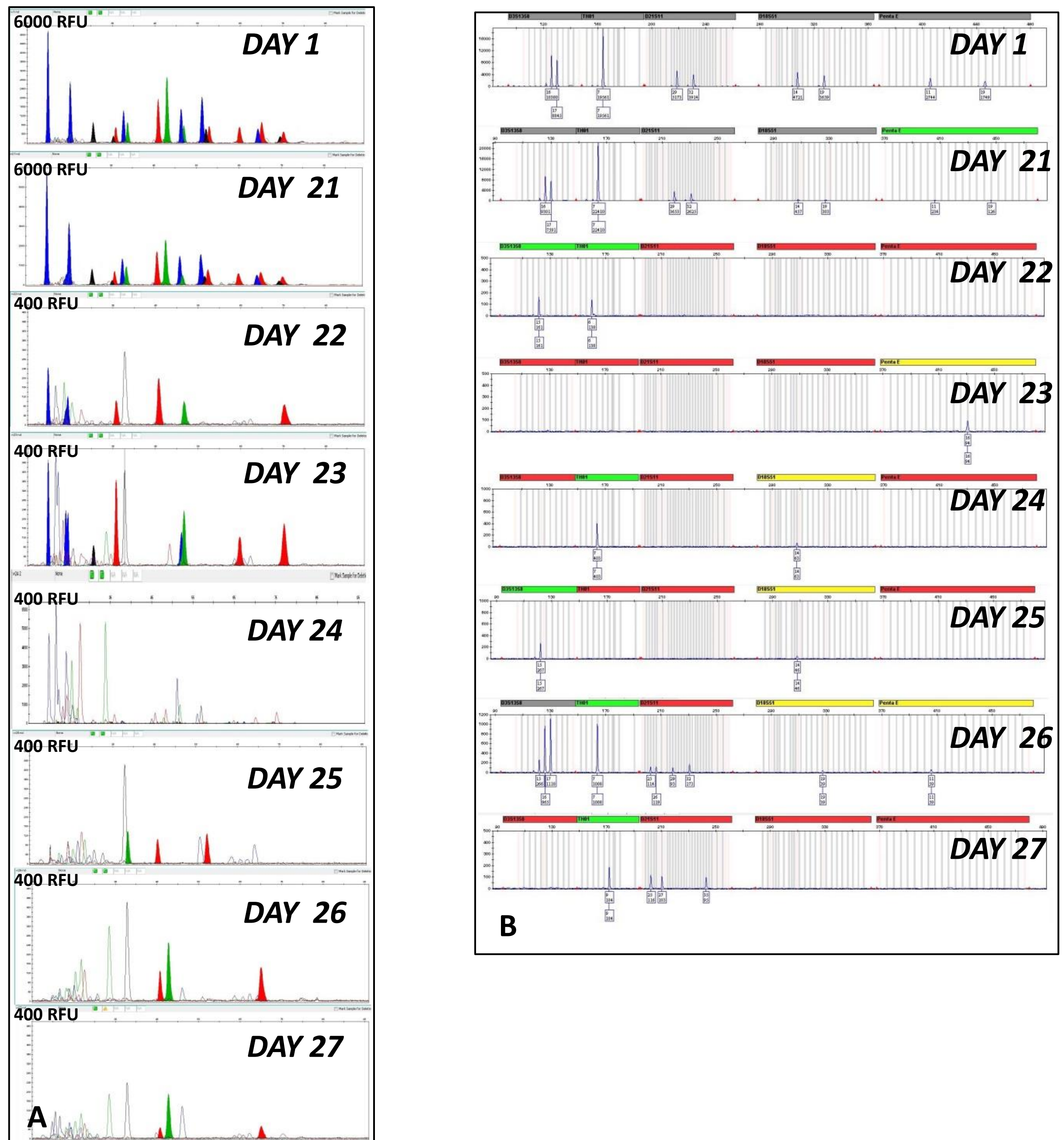


Figure 6-5: A) SNaPshot profiles of 13_{nd} multiplex assay of the simulated environmental samples from Day 1, Day 21 to Day 27, and (B) STR profiles of Powerplex® 16 system of these samples from Day 1, Day 21 to Day 27.

6.3.2.2. 13-plex assays genotyping of the casework samples

Fifty-one crime samples comprising bloodstains on cloths, swabs, and a mat and 2 swabs of trace DNA from 10 crime scenes in Malaysia were profiled after extraction of DNA using a phenol-chloroform method. These samples were subjected to 13-plex assays and Powerplex® 16 System. Results obtained from both analyses were shown as in Table 6-3.

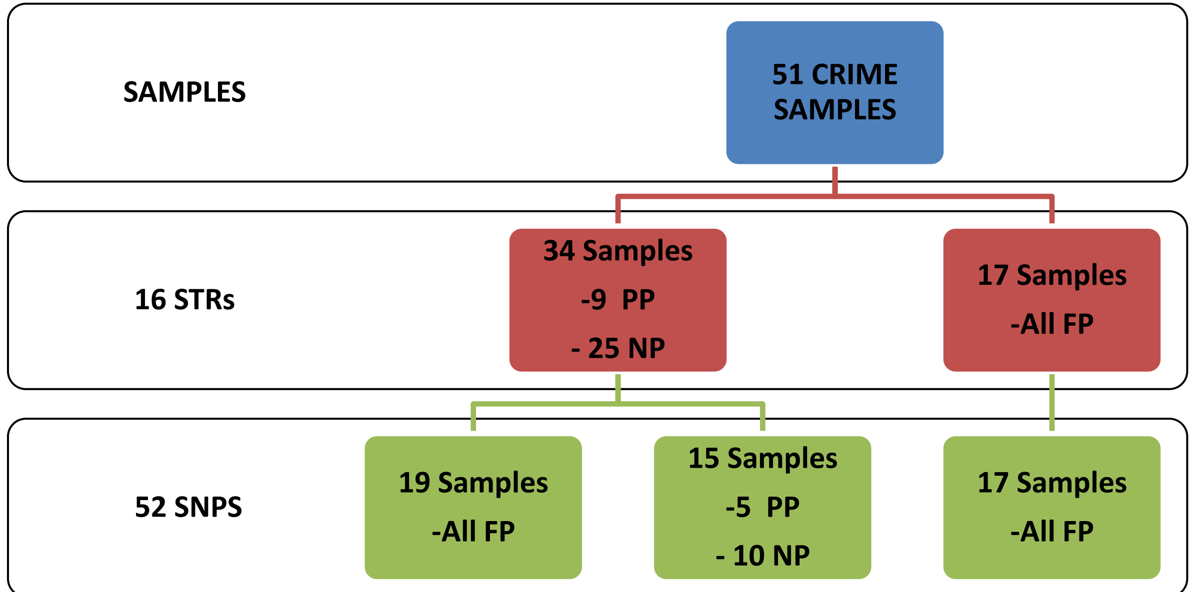
Out of 51 selected samples; 17 full STR profiles were generated using conventional STR analysis; of the remaining 34 samples 9 gave partial profiles and 25 could not be amplified at all. In comparison, with 52 SNP markers, 36 full profiles were successfully amplified from the samples and only 15 were partially amplified or not amplified at all (with 5 partial profiles (with SNPs >26 loci had to be detected to be classed as a partial profile) and no DNA profile was obtained for 10 samples) (Figure 6-6).

Two examples of results obtained from the casework samples are shown in Figure 6-7 and Figure 6-8. In Figure 6-7, full STR and SNPs profiles were generated from the bloodstain on the cloth portion. The DNA concentration was 2.95 ng/µl with IPC_{CT} 28. Whilst in Figure 6-8, no STR loci were amplified but a full SNP profile was produced from this sample. This sample DNA concentration was 15 ng/µl but the IPC_{CT} was undetermined. In this case, the failure of STR loci to be amplified might be because of the presence of inhibitors.

Table 6-3: Results obtained from the analysis of 16 STRs and 52 SNPs on the casework samples.

Case	Samples	In-house Labeled	52 SNPs	Identifiler/Powerplex® Sytem
1	Bloodstains on pillow case	F1	FP	FP
	Bloodstains on bedsheet	F2	FP	PP-13
	Bloodstains on blanket	F3	FP	FP
2	Bloodstains on mat	F4	FP	FP
	Bloodstains on T-shirt	F5	FP	FP
	Bloodstains on pair of jeans	F6	FP	PP-7
	Bloodstains on panties	F7	FP	FP
3	Bloodstains on blouse	F8	FP	PP-8
	Bloodstains on skirt	F9	FP	FP
	Bloodstains on brassiere	F10	FP	PP-8
4	Bloodstains on comforter 1	F11	FP	FP
	Bloodstains on comforter 2	F12	FP	FP
5	Bloodstains on towel	F13	PP-26	PP-8
	Bloodstains on pair of jeans	F14	PP-39	NP
	Bloodstains on blouse	F15	NP	NP
	Bloodstains on camisole	F16	FP	FP
	Bloodstains on skirt	F17	FP	FP
	Bloodstains on sarong	F18	FP	FP
6	Swab of cable 1	F19	NP	NP
	Swab of cable 2	F20	NP	NP
7	Bloodstains on blanket 1	F21	NP	NP
	Bloodstains on blanket 2	F22	FP	FP
	Bloodstains on shirt 1	F23	FP	NP
	Bloodstains on shirt 2	F24	FP	NP
	Bloodstains on shirt 3	F25	FP	NP
	Bloodstains on shirt 4	F26	FP	NP
	Bloodstains on shirt 5	F27	FP	PP-11
	Bloodstains on shirt 6	F28	FP	PP-13
	Bloodstains on shirt 7	F29	FP	FP
	Bloodstains on shirt 8	F30	FP	Weak/Inconclusive
8	Bloodstains on plastic bag 1	F31	PP-50	NP
	Bloodstains on shirt	F32	PP-47	NP
	Bloodstains on pair of jeans	F33	FP	FP
	Bloodstains on towel	F34	FP	FP
	Bloodstains on tissue paper	F35	NP	NP
	Bloodstains on string 1	F36	FP	NP
	Bloodstains on string 2	F37	FP	NP
	Bloodstains on plastic bag 2	F38	NP	NP
9	Bloodstains on pair of short pants	F39	FP	PP-7
	Bloodstains on T-shirt	F40	NP	NP
	Bloodstains on underwear	F41	PP-33	PP-7
	Bloodstains on baton	F42	FP	FP
10	Bloodstains on shirt 1	F43	FP	FP
	Bloodstains on shirt 2	F44	FP	FP
	Bloodstains on pairs of short pants	F45	FP	FP
	Bloodstains on underwear	F46	FP	NP
	Bloodstains on singlet	F47	FP	NP
	Bloodstains on pair of short pants	F48	FP	NP
	Bloodstains on panties	F49	NP	NP
	Bloodstains on brassiere	F50	NP	NP
	Bloodstains on slippers	F51	NP	NP

Notes: FP- Full profile, PP- Partial profile and NP- No profile.



Notes: FP- Full profile, PP- Partial profile and NP- No profile.

Figure 6-6: Number of profiles (STR and SNPs) that have been successfully developed or failed from the crime stains. For SNP, PP means sample developed more than 26 markers.

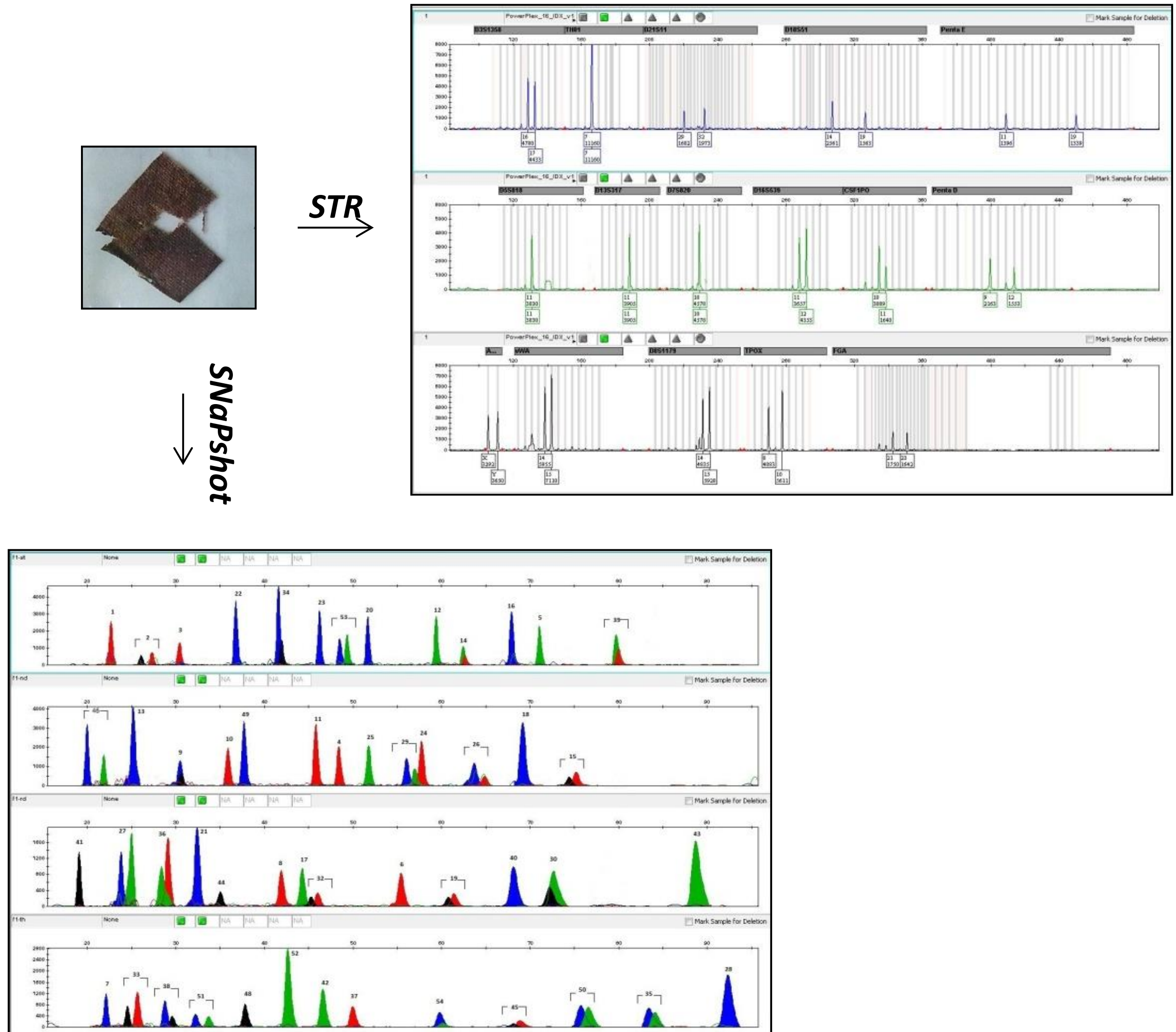


Figure 6-7: Application of SNP and STR assays on the crime sample (with DNA concentration 2.95 ng/µl and IPC_{CT} 28). Full STR (16 markers) and SNaPshot (52 markers) profiles were generated from this sample.

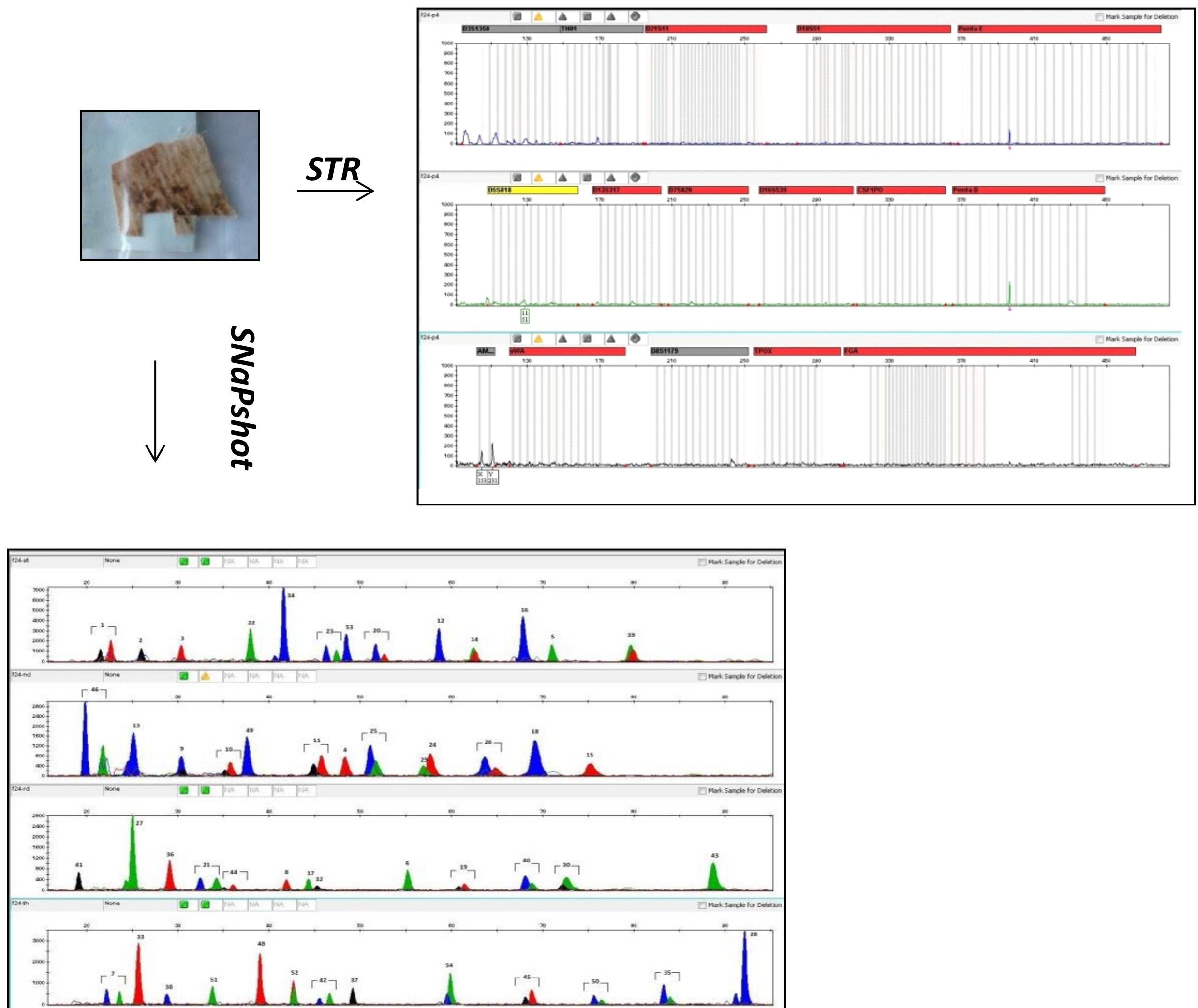


Figure 6-8: Application of SNP and STR assays on the crime sample (with DNA concentration 15 ng/µl and IPC_{CT} undetermined). No STR loci was amplified but a full SNaPshot profile was generated from this sample (a portion of torn shirt bearing bloodstains).

All samples that were successfully amplified using STR or SNP multiplex assays were found to come from the same source. Shown below, in Table 6-4, are the samples that were analysed using DNA.VIEW software. From the results obtained, all full profiles and even partial profiles gave higher match probability with theta (θ) corrections (0.01 and 0.03) is incorporated. This shows the SNP multiplex assays have potential to be used in forensic casework analysis.

Table 6-4: Statistical evaluation of SNP data using DNA.VIEW software.

Case	Samples	In-house Labeled	Source of the samples	52 SNPs	Theta (θ)	
					0.01	0.03
1	Bloodstains on pillow case	F1	Samples are from the same source.	FP	$1e^{18}$	$720e^{15}$
	Bloodstains on bedsheet	F2		FP		
	Bloodstains on blanket	F3		FP		
2	Bloodstains on mat	F4	Samples are from the same source.	FP	$23e^{18}$	$15e^{18}$
	Bloodstains on T-shirt	F5		FP		
	Bloodstains on pair of jeans	F6		FP		
	Bloodstains on panties	F7		FP		
3	Bloodstains on blouse	F8	Samples are from the same source.	FP	$12e^{18}$	$8.6e^{18}$
	Bloodstains on skirt	F9		FP		
	Bloodstains on brassiere	F10		FP		
4	Bloodstains on comforter 1	F11	Samples are from the same source.	FP	$2.7e^{18}$	$1.9e^{18}$
	Bloodstains on comforter 2	F12		FP		
5	Bloodstains on towel	F13	Samples are from the same source.	PP-26	$1.7e^{12}$	$1.2e^{12}$
	Bloodstains on pair of jeans	F14		PP-39	$17e^{15}$	$11e^{15}$
	Bloodstains on blouse	F15		NP		
	Bloodstains on camisole	F16	Samples are from the same source.	FP	$3.2e^{21}$	$1.9e^{21}$
	Bloodstains on skirt	F17		FP		
	Bloodstains on sarong	F18		FP		
6	Swab of cable 1	F19		NP		
	Swab of cable 2	F20		NP		
7	Bloodstains on blanket 1	F21		NP		
	Bloodstains on blanket 2	F22	Samples are from the same source.	FP	$200e^{21}$	$110e^{21}$
	Bloodstains on shirt 1	F23		FP		
	Bloodstains on shirt 2	F24		FP		
	Bloodstains on shirt 3	F25		FP		
	Bloodstains on shirt 4	F26		FP		
	Bloodstains on shirt 5	F27		FP		
	Bloodstains on shirt 6	F28		FP		
	Bloodstains on shirt 7	F29		FP		
	Bloodstains on shirt 8	F30		FP		
8	Bloodstains on plastic bag 1	F31	Samples are from the same source.	PP-50	$41e^{18}$	$28e^{18}$
	Bloodstains on shirt	F32		PP-47	$51e^{18}$	$30e^{18}$
	Bloodstains on pair of jeans	F33		FP	$32e^{18}$	$23e^{18}$
	Bloodstains on towel	F34		FP		
	Bloodstains on tissue paper	F35		NP		
	Bloodstains on string 1	F36	Samples are from the same source.	FP	$32e^{18}$	$23e^{18}$
	Bloodstains on string 2	F37		FP		
	Bloodstains on plastic bag 2	F38		NP		
9	Bloodstains on pair of short pants	F39	Samples are from the same source.	FP	$460e^{18}$	$280e^{18}$
	Bloodstains on T-shirt	F40		NP		
	Bloodstains on underwear	F41	Samples are from the same source.	PP-33	$28e^{9}$	$23e^{9}$
	Bloodstains on baton	F42		FP	$460e^{18}$	$280e^{18}$
10	Bloodstains on shirt 1	F43		FP	$570e^{18}$	$360e^{18}$
	Bloodstains on shirt 2	F44		FP		
	Bloodstains on pairs of short pants	F45		FP		
	Bloodstains on underwear	F46		FP		
	Bloodstains on singlet	F47		FP		
	Bloodstains on pair of short pants	F48		FP		
	Bloodstains on panties	F49		NP		
	Bloodstains on brassiere	F50		NP		
	Bloodstains on slippers	F51		NP		

6.4. DISCUSSION

In the case of natural and mass disasters, missing person's cases and forensic casework, highly degraded biological samples are often encountered. Conventional methods of identification such as fingerprint, forensic anthropology and dental matching can be inadequate when remains are highly fragmented and decomposed. DNA typing using industry standard short tandem repeats (STRs) often becomes the principal means of identification (Brenner 2006, Hughes-Stamm et al. 2011).

In this study 13-plex assays were applied to the simulated crime samples and paternity samples, in order to evaluate the usefulness of these SNP markers in handling problematic samples. As reported by Stamm and colleagues (2011), their results indicated mini-STRs and SNP markers are usually more successful in typing degraded samples and in cases of extreme DNA degradation (≤200 bp) and template amounts below 250 pg, mini-STR and SNP analysis yielded significantly more complete profiles and lower match probabilities than corresponding STR profiles (Hughes-Stamm et al. 2011).

6.4.1. Section One: Kinship studies

One of the main aims of this chapter was to study the identification ability of these SNP markers. Therefore the first approach was applied these 13-plex assays on the paternity samples having different scenarios. Results obtained from this study showed that normal trio paternity cases gave higher CPI and probability % than normal duo cases. In addition, STR analysis also gave higher CPI and probability % than SNP analysis. However, in cases where STR mutation(s) were present SNP analysis gave higher CPI and probability % than STRs. This phenomenon could be because in commonly used STR, one mutation is expected to occur in one of the loci typed in approximately 3% of all meioses. If one or two genetic inconsistencies are detected between a child and the alleged parents may not be the true biological parent. In contrast, mutations occur 100,000 times less frequently in SNPs than in STRs, and inconsistencies caused by mutations in SNP loci are extremely rare (Borsting et al. 2012a).

6.4.2. Section Two: The analysis of forensic biological samples

6.4.2.1. Simulated casework samples

Poor sample handling and storage may lead to further loss or degradation of samples and could impact obtaining useful results. Additional factors that may compromise sample integrity include high humidity, temperature, persistence of nucleases and other chemical agents as well as other sub-optimal conditions that may occur not only during transport, but also within storage facilities (Lee et al. 2012).

In order to test whether these 13-plex assays will be suitable to be used on the Malaysian crime samples, simulated samples were exposed to typical Malaysian weather, very high humidity and temperature for 27 days. The extent of DNA degradation and inhibition can be seen as shown in Figure 6-4. Samples from Day 21 to Day 27 showed the presence of fungi on the bloodstain spot and on the cloth itself. The results obtained from STR and SNP analysis also demonstrate that samples from Day 22 to Day 27 failed to amplify.

Spontaneous degradation by hydrolysis and oxidation will further modify DNA structure at a much slower speed. However, a fundamental knowledge of the types of DNA damage encountered in forensic stains that would lead to amplification failure has not been yet been provided. There are a few other factors that might lead to DNA degradation through environmental conditions or microbial activity such as radiation and oxidation processes, the availability of sufficient nutrients, pH and ionic concentration for the growth of microorganisms (Alaeddini et al. 2010).

6.4.2.2. Real casework samples

DNA becomes progressively more fragmented as biological tissue degrades resulting in a decreasing ability to gain a complete DNA profile. Successful identification of samples exhibiting very high levels of DNA degradation may be complicated by presenting in minute quantities. The industry standard method for human DNA identification utilizing short tandem repeats (STR) may produce partial or no DNA profiles with such samples (Hughes-Stamm et al. 2011).

In this study, the results obtained demonstrate that in some instances the SNPs can generate full profiles from DNA extracts that yielded no or partial STR loci. The samples that were successfully typed have input DNA from undetermined up to 100 ng (quantified using the Quantifiler® kit), and this suggests that some of the samples were highly degraded or inhibited as STR profiles would have otherwise been expected.

The DNAVIEW calculation shows that even the partial profiles such as 26 SNP markers and above gave a high power of discrimination. Overall, out of 51 crime samples, STR typing was able to amplify 17 full profiles and 9 partial profiles. By contrast, with SNP analysis, 36 full profiles and 5 partial profiles had been successfully amplified from the DNA extracts.

Sanchez and colleagues (2006) reported that the full set of 52 SNP markers were successfully amplified from partially degraded DNA samples using 200 to 500 pg DNA input (Sanchez et al. 2006a). Børsting and colleagues (2011) have also reported that in their study, complete SNP profiles were obtained in 80% of the highly degraded samples and at least 41 SNPs were typed in the last 20% (Børsting et al. 2011).

The high sensitivity and the short amplicon sizes make the assay very suitable for typing of partially degraded DNA, and the low mutation rate of SNPs makes the assay very useful for relationship testing. Indirectly, these advantages make the assay well suited for disaster victim identifications, where the DNA from the victims may be highly degraded and the victims are identified via investigation of their relatives (Børsting et al. 2013).

CHAPTER 7

GENERAL DISCUSSION AND FUTURE WORK

GENERAL DISCUSSION AND FUTURE WORK

7.1. GENERAL DISCUSSION

SNPs have already proved to be valuable markers and SNPs are used by some laboratories in routine case work investigations. SNPs are mainly used as supplementary markers, but there are situations where SNPs may be the best choice of markers. If the DNA is highly degraded, there is a much higher chance of obtaining results with SNP markers than with short tandem repeat (STR) markers, because SNPs may be amplified on very short amplicons (<100 bp) (Sanchez et al. 2006a, Butler et al. 2008, Phillips et al. 2008b)

The performance of the 52plex SNP assay was recently compared to the use of STRs and VNTRs (variable number of tandem repeats) in paternity casework. The typical paternity indices (PIs) obtained based on the 52 SNPs were 5–50 times lower than the typical PIs based on 15 STRs or 7 VNTRs. However, the mutation rate of SNPs is expected to be approximately 100,000 times lower than the mutation rate of tandem repeats, and this makes the SNPforID 52plex SNP assay a valuable investigative tool in forensic casework where potential relatives are involved. This is especially true in cases where various family scenarios are possible or in cases where a possible mutation has been detected in one of the tandem repeat loci, and supplementary investigations are required. Furthermore, the SNPforID 52plex SNP assay may be highly useful in disaster victim investigations, because the assay can generate complete profiles from highly degraded DNA, where the commercial STR kits fail (Borsting et al. 2009).

In developing multiplex assays, 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 analyze the samples.

Few studies had been carried out involved the PCR/SBE primers, to adjust and obtain balanced SNP profile. In this study, four sets of 13-plex assays (*13_{st}, 13_{nd}, 13_{rd} and 13_{th}*) have been developed from the 52 SNP markers identified (Sanchez et al. 2006a) and the single base primer extension method has been chosen to be used with the SNaPshot Multiplex kit. The main reason for the reduction of SNP markers from the original study is for easier interpretation and to accommodate the ABI 310 capability in detecting SNP alleles. These assays will also be applied to the forensic casework samples, therefore an easy and reliable tool is needed to interpret the SNP data.

Any technology that is applied to forensic casework, such as CE systems must be rigorously evaluated through a comprehensive validation program (Moretti et al. 2001, Butler et al. 2004). The validation experiments reveal the operational parameters that are employed in the assessment of peaks detected during CE analysis (Butler et al. 2004).

In comparing the reproducibility of 13-plex assays generated from both genetic analyzers, a concordance study had been carried out using polymer, POP-7™ and 3500 genetic analyzer, but the resolution of the profiles were found to be poor. Further experiments were undertaken and as a result, POP-6™ was decided to be the best polymer to be used in detecting SNP alleles using 3500 genetic analyzer.

All these observations were very important because it gives new insight in SNP genotyping using the 3500 analyzer especially for profiles that would have many heterozygous alleles. The CE instruments should have a high resolving capability with high precision to be able to identify alleles that have one or two bases in size difference (Isenberg et al. 1998).

Apart from different software and capabilities between ABI 310 and ABI 3500, overall, the four sets of 13-plex assays introduced in the study were reliable and could be used in the SNP genotyping of the population and casework samples. In fact, the sensitivity of these assays on both genetic analyzers was as low as 30 pg. In some profiles with only 30 pg input, the samples showed one or two allele drop-outs, especially in the

heterozygous alleles. Therefore, the best SNP profile obtained with full amplification of 52 alleles was at DNA input more than 60 pg. Full SNP profiles were also successfully amplified from DNA samples of more than 70 ng with no allele drop-out but the imbalanced peak heights were observed. Overall, as for forensic casework, these assays (*13_{st}, 13_{nd}, 13_{rd} and 13_{th}*) were reliable and could be useful tools in genotyping low template DNA less than 200 pg, and particularly degraded DNA.

In forensic laboratories, an amount of DNA below a specific level (between 100 and 200 pg) has been variously defined as low template DNA (Gill & Buckleton 2009). These 13-plex assays provide useful and reliable tools in detecting alleles from samples with input DNA less than 30 pg or higher than 70 ng.

Besides that, although these 52 SNP markers were meant to be used in forensic human identification and not ethnicity prediction, the results obtained using these 52 SNP markers could possibly differentiate between the Indians and the other two ethnic groups.

The results obtained from this 52 SNPs study combined with those previously published suggest that a complex model of admixture could have occurred between the Malays, the Chinese and the Indians or even with the indigenous people. This is consistent with the hypothesis that human populations are not discrete groups, since admixture commonly occurs between neighbouring populations (Sanchez et al. 2006a, Serre & Paabo 2004). Therefore, it is necessary to take into account for the future studies especially regarding different ancestry origin among the populations.

This study not only revealed least differences, F_{ST} for most markers between the three studied ethnic populations, the Malays, the Chinese and the Indians that indicates the closeness of these three ethnic groups, but also shows that the 52 SNP markers can provide a high discrimination capacity which is essential for forensic applications.

Limitations/Challenges of SNP Analysis

There are a few limitations of SNP analysis was identified to be used as core forensic technology such as no commercial kits available, no widely established core loci, having mixture resolution issues/interpretation, large multiplexing assays are required, linkage studies very limited and substructure due to low mutation rate and there are multiple typing platforms that make universal SNP selection difficult (Butler et. al.2008).

7.2. FUTURE WORK

As for future work, marker sets that are based on small insertion/deletion (INDEL) alleles can serve as useful supplementary or stand-alone assays for human identification. The presence or absence of a small insertion or deletion (INDEL), technically a SNP, offers the value of a more simplified analytical process. In addition, the instrumentation for INDEL analyses is commonly found in forensic laboratories and essentially, INDELs perform analytically similar to that of STRs (Larue et al. 2012, Pereira, Gusmao 2012).

Further population studies on variation in individual admixture proportions will be interesting for Malaysian's multi-ethnic populations. Current interest that growing is establishing SNP tests that are able to predict hair, skin and eye colour variation (Kayser & Schneider 2009, Walsh et al. 2013) in populations, along with the ancestry informative markers (AIM) (Li et al. 2008, Fondevila et al. 2013, Pereira et al. 2012, Wall et al. 2013) 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 .

APPENDICES

APPENDIX A: Approval letter of samples collection from Department of Chemistry Malaysia, Ministry of Science Technology and Innovation, Malaysia.

JABATAN KIMIA MALAYSIA
KEMENTERIAN SAINS, TEKNOLOGI DAN INOVASI
DEPARTMENT OF CHEMISTRY MALAYSIA
MINISTRY OF SCIENCE, TECHNOLOGY AND INNOVATION

22nd April 2010

Reader
Forensic Genetics
School of Forensic and Investigate Sciences
University of Central Lancashire
Preston PR1 2HE
Lancashire
United Kingdom

(Attn: Mr. William Goodwin)

Dear Sir,

RE: PROVISION OF BIOLOGICAL SAMPLES FOR THE RESEARCH PROJECT OF SHARIZAH ALIMAT

With reference to the request for biological samples, I am pleased to inform you that we, the Department of Chemistry Malaysia, are happy to provide the required samples, namely: bloodstains on FTA card for the three major ethnic groups of Malaysia (Indian, Malay, and Chinese). These are from a databank held by the Department of Chemistry and were collected from volunteers from patients at the General Hospital, Kuala Lumpur. The samples were collected with informed consent and have been held for the purpose of generating reference databases for DNA markers since the late 1990s. The bloodstains will be provided as anonymous samples, with details only related to sex and ethnic group provided. We will provide approximately 60 samples of each ethnic group.

In addition, we also agree to give Sharizah access to casework samples, in the form of remaining stains only after we have completed our analysis. We will allow Sharizah to sample the remaining stains and transport material to Preston, for the purpose of carrying out her PhD research.

Thank You.

Yours Sincerely,

(Lim Kong Boon)
Head of Serology/DNA Section
Forensic Division
Department Of Chemistry Malaysia
kblim@kimia.gov.my
f Director of Forensic Division

c.c Sharizah Alimat

Jalan Sultan, 46661 Petaling Jaya, Selangor Darul Ehsan
Tel.: 603 - 7985 3000 Fax: 603 - 7985 3092
www.kimia.gov.my

APPENDIX B: Approval notification from Health, Safety and Ethics Committee of the School of Forensic and Investigative Sciences.

Proposal Decision Form

Health, Safety and Ethics Committee
School of Forensic and Investigative Sciences

Reference No: 053

The proposal entitled Application of single nucleotide polymorphisms (SNPs to forensic casework in Malaysia

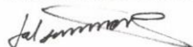
and submitted by Sharizah Alimat was considered by the Health, Safety and Ethics Committee of the School of Forensic and Investigative Sciences.

The committee's decision regarding the proposal was:

	Approved	You may proceed with the research project
x	Approved by Chair's Action	You may now proceed with the research project
	Approved Pending Minor Revisions (see comments below)	You must re-submit the proposal according to the specifications; once you have done this, the committee chairperson will review the revision and notify you that you may proceed
	Requires Major Revision (see comments below)	The proposal must be revised extensively and resubmitted to the committee as a whole
	Rejected (see comments below)	The research proposal is not acceptable

Comments:
Approved – pending receipt of correspondence indicating that it is ok to use the database, in particular unknown samples. **17/5/10 - Correspondence (hard copy) has been received and filed, and approval has been given.**

The committee thanks you for your submission and wishes you all the best,



Tal Simmons, PhD, DABFA, RFP
Chair, Health, Safety and Ethics Committee

APPENDIX C: Records of meteorological data.

Year	Month	Day	Temperature (°C)		Rainfall (mm)	Relative Humidity (%)	
			Maximum	Minimum		8.00 a.m	2.00 p.m
2011	08	08	33.9	23.9	0.0	92	61
2011	08	09	33.6	24.5	0.0	92	58
2011	08	10	31.8	23.7	0.0	92	86
2011	08	11	34.2	22.8	0.0	91	53
2011	08	12	33.3	24.8	12.1	89	55
2011	08	13	33.4	22.7	13.4	91	59
2011	08	14	32.0	23.4	1.2	92	70
2011	08	15	33.2	23.4	0.0	96	61
2011	08	16	33.0	23.2	0.0	93	58
2011	08	17	33.9	23.5	0.0	91	51
2011	08	18	33.9	24.6	0.0	84	51
2011	08	19	34.6	24.2	3.7	92	53
2011	08	20	30.2	23.4	16.3	91	88
2011	08	21	33.7	24.4	13.5	92	53
2011	08	22	33.0	23.9	2.0	92	90
2011	08	23	32.6	23.4	13.1	92	81
2011	08	24	32.4	24.0	1.2	88	66
2011	08	25	32.5	23.3	0.0	96	83
2011	08	26	30.2	23.4	29.4	92	93
2011	08	27	32.4	23.9	0.0	88	67
2011	08	28	31.0	24.8	17.0	92	69
2011	08	29	30.6	23.0	0.0	96	64
2011	08	30	33.1	24.0	0.0	87	56
2011	08	31	31.1	25.5	0.0	89	66
2011	09	01	33.4	22.4	0.0	87	52
2011	09	02	31.4	24.7	0.0	92	75
2011	09	03	32.6	23.4	0.0	85	58

APPENDIX D: Real forensic casework samples used in the study.

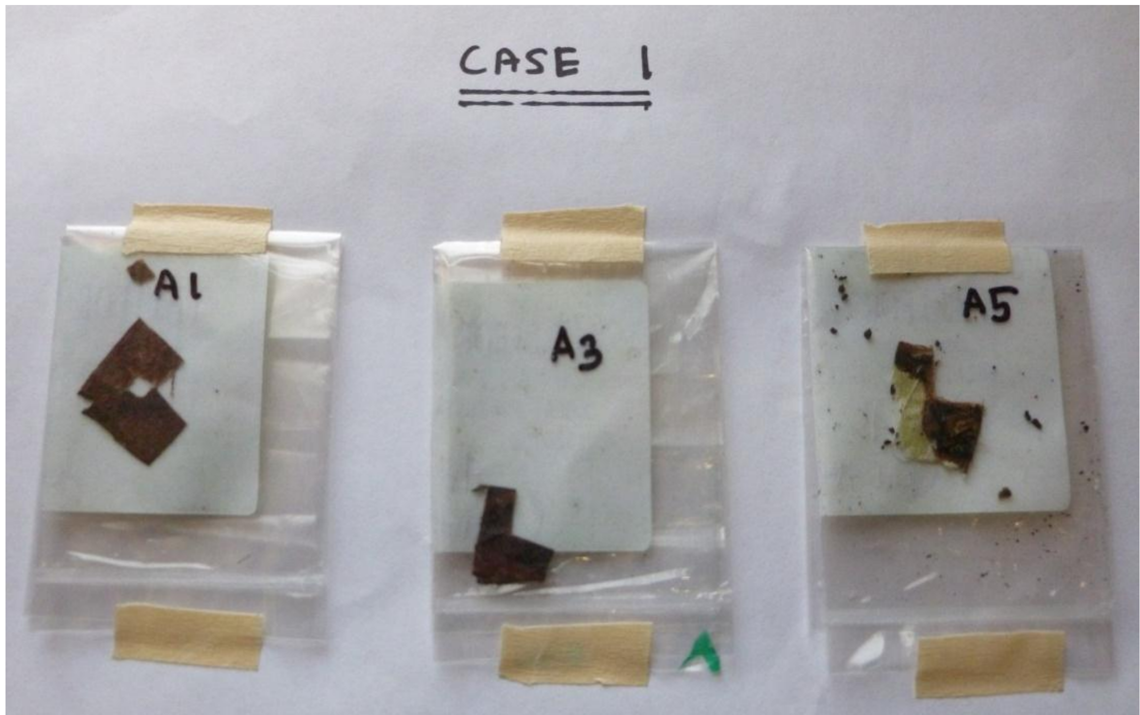


Figure 1: Shown above are the casework samples from Case 1.

Figure 2: Shown above are the casework samples from Case 2.

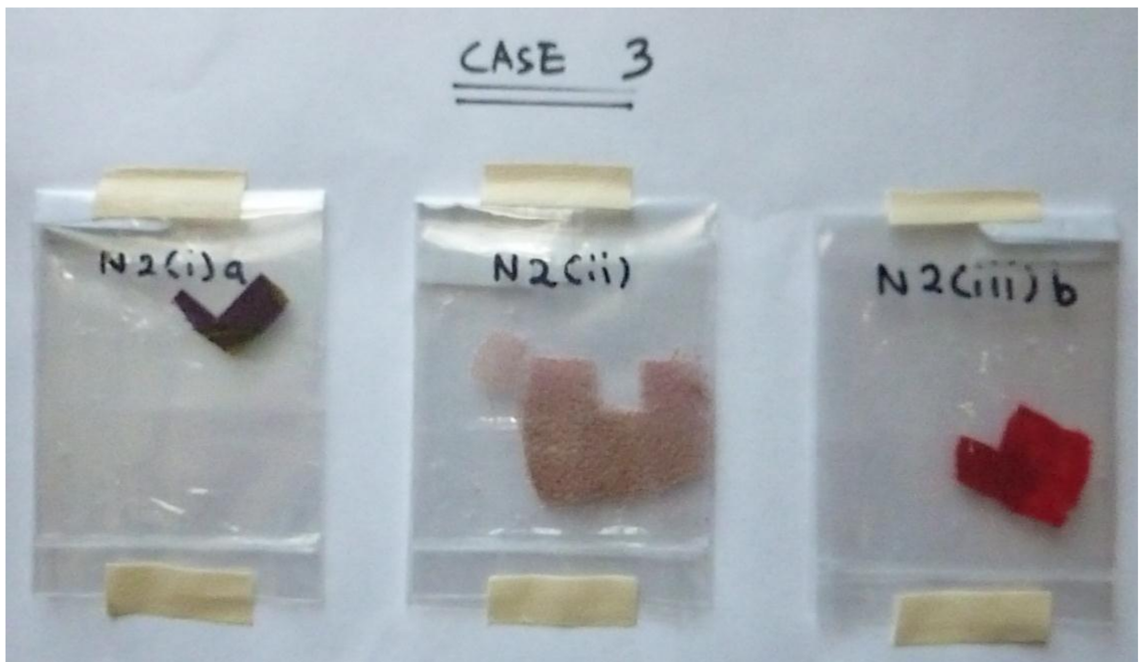


Figure 3: Shown above are the casework samples from Case 3.

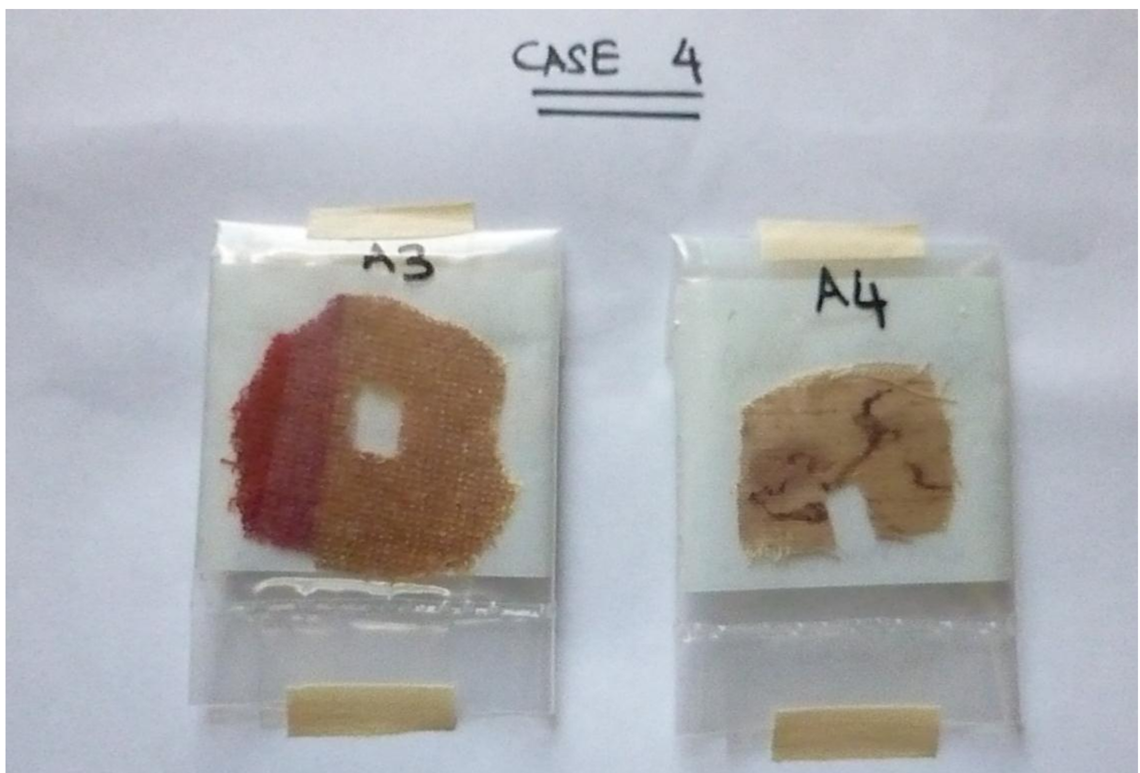


Figure 4: Shown above are the casework samples from Case 4.

Figure 5: Shown above are the casework samples from Case 5.

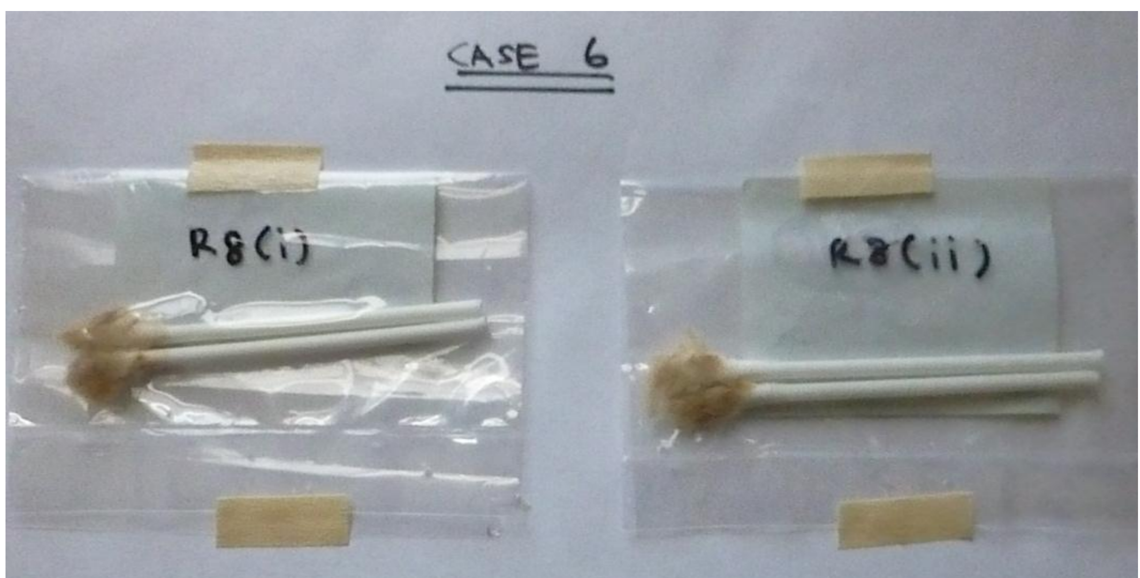


Figure 6: Shown above are the casework samples from Case 6.

Figure 7: Shown above are the casework samples from Case 7.

Figure 8: Shown above are the casework samples from Case 8.

Figure 9: Shown above are the casework samples from Case 9.

Figure 10: Shown above are the casework samples from Case 10.

APPENDIX E: Documents set-up for Allelic Discrimination study

Figure 11: An example of a Pre-Read document created.

Figure 12: Detector was created and selected for targeted SNP markers in a Pre-Read document.

.

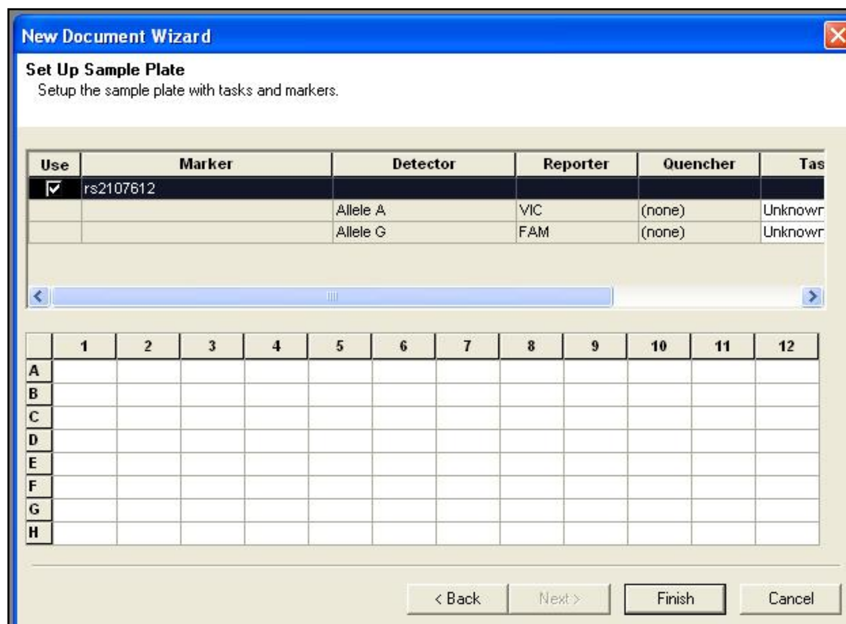


Figure 13: The sample plate was set-up with tasks and markers chosen in a Pre-Read document.

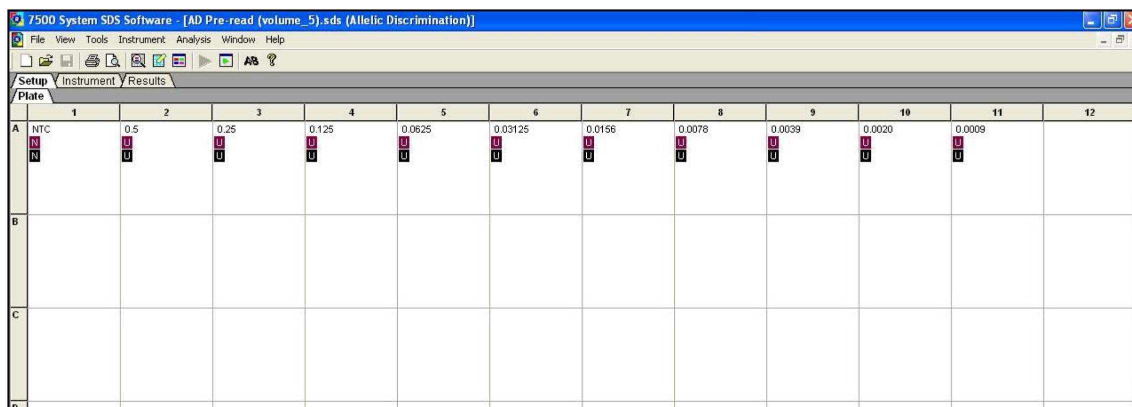


Figure 14: An example of the Pre-Read template.

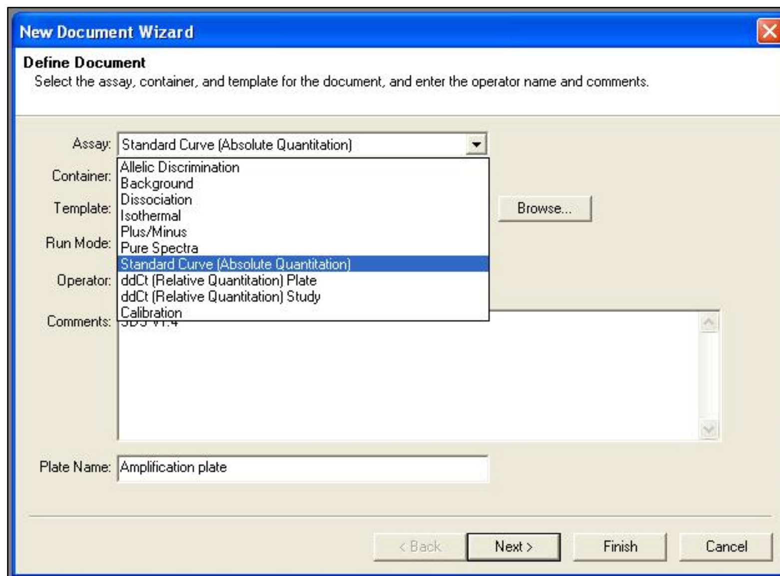


Figure 15: Selecting the assay type to create the Amplification document

Figure 16: An example of the Amplification template with sample names, tasks, a specific marker and detector.

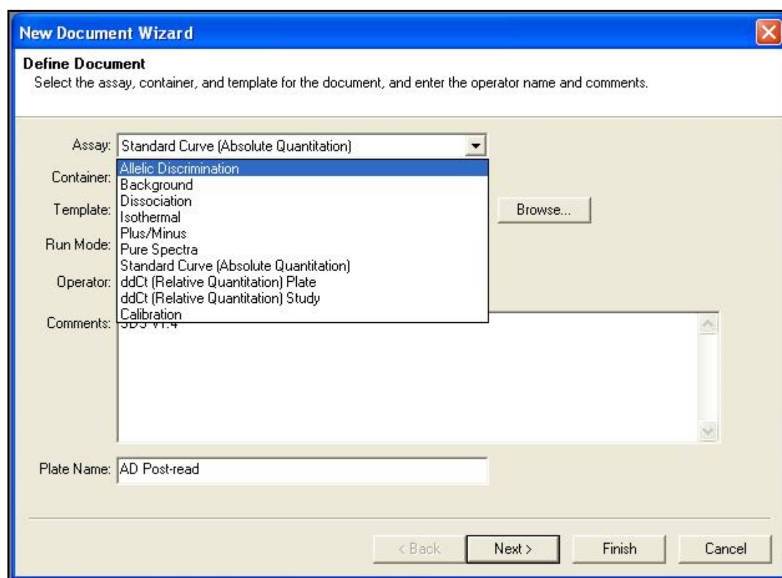


Figure 17: Creating an AD Post-Read document.

Figure 18: An example of a Post-Read document with sample names, tasks, a marker and detector of targeted SNP.

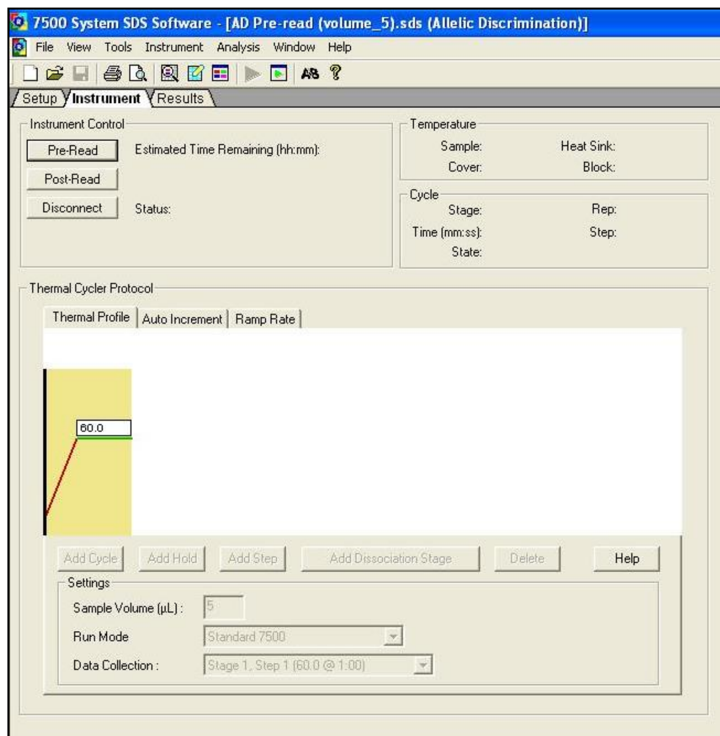


Figure 19: Instrument conditions for a Pre-Read analysis.

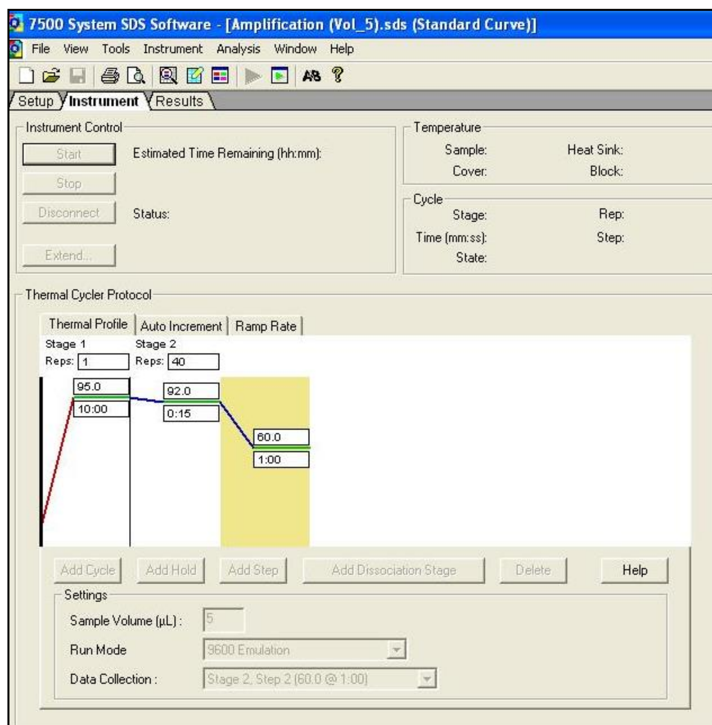


Figure 20: Instrument conditions for an Amplification analysis.

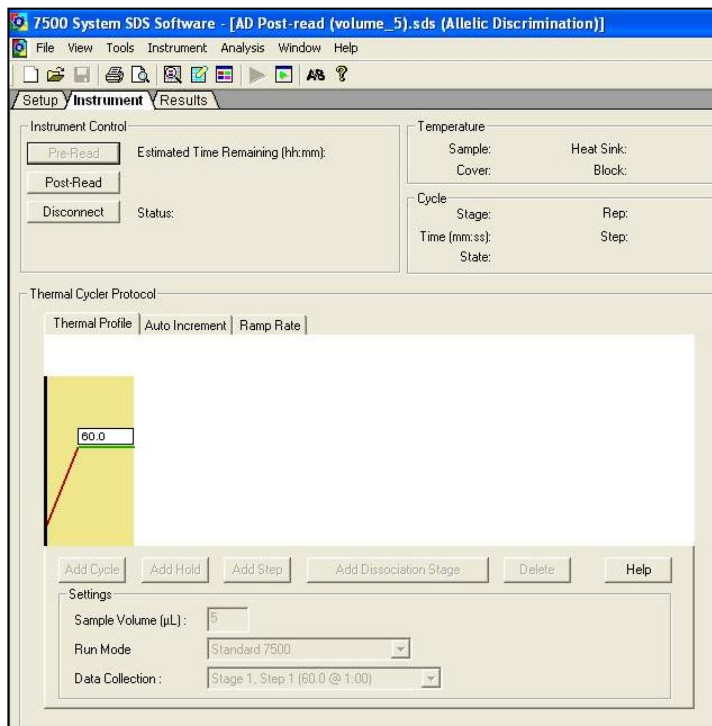


Figure 21: Instrument conditions for a Post-Read analysis.

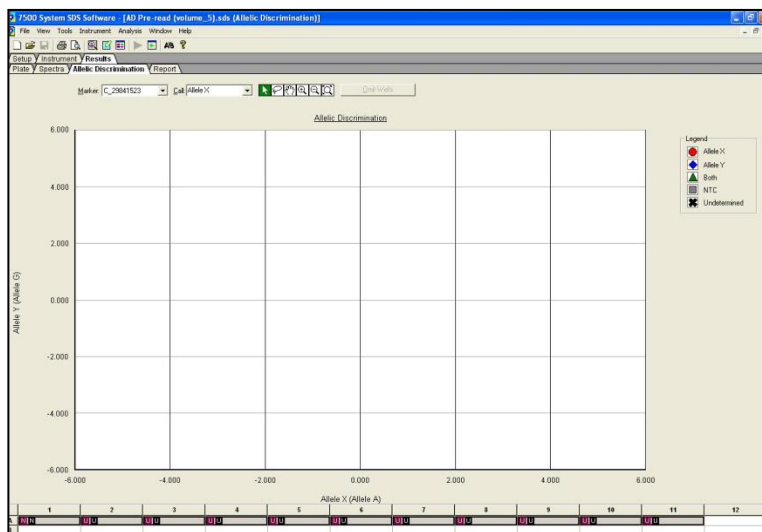


Figure 22: No allele was observed in the Allelic Discrimination tab.

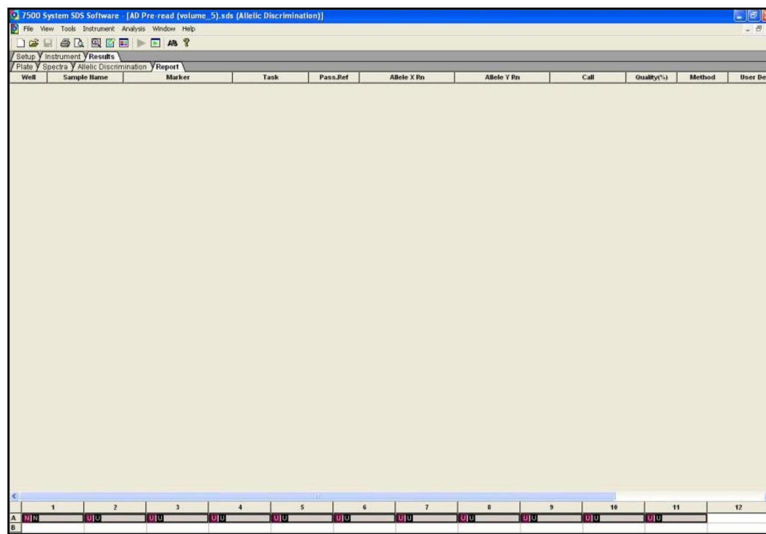


Figure 23: No report of analysis was generated.

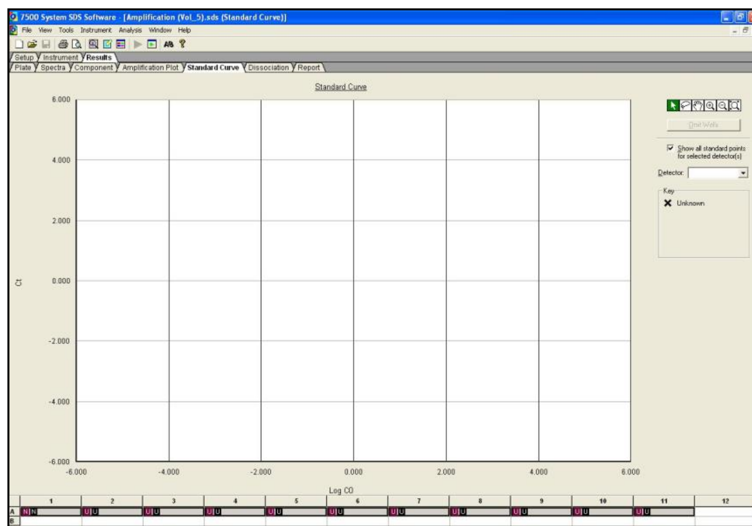


Figure 24: The Allelic Discrimination tab of the AQ analysis.

APPENDIX F: DNA.VIEW calculation of 52 SNP markers for tested samples.

Case	1		Scenario	1				
	M		Mother	1000-0097				
	C		Child	1000-0098				
	F		Tested Man	1000-0099				
C: M + F								
/C: M+ ?								
	Malay		cumulative LR	2.16363e6	Posterior probability=	99.99995%	assuming prior=	50%
►	Malay					M	C	F
	A01	SNP	2.32979	1 / c	c=0.429	C	C	C
	A02	SNP	1.54225	1 / c	c=0.648	C	C	C
	A03	SNP	1	1 / (a+t)	a=0.164 t=0.836	AT	AT	T
	A04	SNP	1	1 / (a+t)	a=0.564 t=0.436	A T	AT	A
	A05	SNP	1.86441	1 / 2g	g=0.268	A	G A	GA
	A06	SNP	1.46667	1 / 2a	a=0.341	A	A	AT
	A07	SNP	2.2	1 / 2a	a=0.227	G A	A	G A
	A08	SNP	2.31579	1 / c	c=0.432	T	CT	T
	A09	SNP	0.873016	1 / 2c	c=0.573	G	GC	G C
	A10	SNP	1	1 /(c+t)	c=0.182 t=0.818	CT	C T	T
	A11	SNP	1.92982	1 / 2c	c=0.259	C T	C	C T
	A12	SNP	0.70677	1 / 2a	a=0.705	A	A	G A
	A13	SNP	1	1 / (a+g)	g=0.855 a=0.145	G A	G A	G A
	A14	SNP	1.77419	1 / a	a=0.564	T	AT	A
	A15	SNP	1	1 / (c+t)	c=0.35 t=0.65	C T	C T	C T
	A16	SNP	1	1 / (g+t)	g=0.841 t=0.159	GT	G T	G
	A17	SNP	3.66667	1 / 2g	g=0.136	A	GA	GA
	A18	SNP	2.5	1 / t	t=0.4	G	G T	T
	A19	SNP	1.5	1 / t	t=0.667	T	T	T
	A20	SNP	1.07843	1 / 2g	g=0.464	T	G T	GT
	A21	SNP	1.71094	1 / a	a=0.584	A	A	A
	A22	SNP	1.80328	1 / 2a	a=0.277	GA	A	GA
	A23	SNP	0.628571	1 / 2g	g=0.795	G	G	GA
	A24	SNP	1.17021	1 / 2c	c=0.427	CT	C	CT
	A25	SNP	1.19022	1 / a	a=0.84	A	A	A
	A26	SNP	1.20879	1 / 2g	g=0.414	T	G T	G T
	A27	SNP	1.65414	1 / a	a=0.605	G	G A	A
	A28	SNP	1	1 / (a+g)	g=0.832	GA	GA	G
	A29	SNP	0.555556	1 / 2g	g=0.9	G	G	GA
	A30	SNP	1	1 / (a+c)	a=0.541 c=0.459	AC	A C	AC
	A32	SNP	0.866142	1 / 2t	t=0.577	C	C T	CT
	A33	SNP	1.74603	1 / 2c	c=0.286	T	C T	CT
	A34	SNP	1	1 / (c+g)	g=0.541 c=0.459	GC	G C	G
	A36	SNP	1.34146	1 / t	t=0.745	AT	T	T
	A37	SNP	1.52778	1 / 2c	c=0.327	T	C T	CT
	A38	SNP	1.06796	1 / 2c	c=0.47	C	C	GC
	A39	SNP	1.22222	1 / 2a	a=0.409	A	A	A T
	A40	SNP	3.05556	1 / a	a=0.327	GA	A	A
	A41	SNP	0.601093	1 / 2c	c=0.832	C	C	C T
	A42	SNP	0.833333	1 / 2g	g=0.6	A	GA	GA
	A43	SNP	0.578947	1 / 2a	a=0.864	A	A	G A
	A44	SNP	1.69231	1 / 2t	t=0.295	CT	T	CT
	A45	SNP	1.3949	1 / c	c=0.717	C	C	C
	A46	SNP	4.58333	1 / a	a=0.218	G A	A	A
	A48	SNP	2.97297	1 / c	c=0.336	T	C T	C
	A49	SNP	1	1 / (a+g)	g=0.804 a=0.2	GA	GA	GA
	A50	SNP	2.36559	1 / a	a=0.423	G A	A	A
	A51	SNP	1	1 / (a+g)	g=0.527 a=0.473	G A	G A	A
	A52	SNP	0.700637	1 / 2a	a=0.714	AT	A	A T
	A54	SNP	1.05769	1 / 2g	g=0.473	G	G	G A
	A35	SNP	2.01835	1 / a	a=0.495	GA	A	A
	A53	SNP	1.12245	1 / 2g	g=0.445	G	G	G A
	cumulative LR		2.16363e6					

Case	2	Scenario	2
	M	Mother	1000-0100
	C	Child	1000-0101

C: M + ?
/C: ?+ ?

Malay	cumulative LR	796.423	Posterior probability=	99.9%	assuming prior=	50%

	Malay					M	C
►	A01	SNP	0.8	1 / 2t	t=0.573	T	C T
	A02	SNP	0.774648	1 / 2c	c=0.645	C	CT
	A03	SNP	1.19022	1 / t	t=0.84	T	T
	A04	SNP	0.887097	1 / 2a	a=0.564	A	AT
	A05	SNP	0.68323	1 / 2a	a=0.732	A	G A
	A06	SNP	0.758621	1 / 2t	t=0.659	A T	T
	A07	SNP	1.28824	1 / g	g=0.776	G	G
	A08	SNP	0.88	1 /2 t	t=0.568	T	CT
	A09	SNP	1.7381	1 / c	c=0.575	C	C
	A10	SNP	2.75	1 /2c	c=0.182	CT	CT
	A11	SNP	1.34356	1 / t	t=0.744	T	T
	A12	SNP	1.4129	1 / a	a=0.708	A	A
	A13	SNP	0.585106	1 / 2g	g=0.855	G	G A
	A14	SNP	0.887097	1 / 2a	a=0.564	A	AT
	A15	SNP	1.53147	1 / t	t=0.653	T	T
	A16	SNP	1.18378	1 / g	g=0.845	G	G
	A17	SNP	1.15263	1 / a	a=0.868	A	A
	A18	SNP	2.48864	1 / t	t=0.402	T	T
	A19	SNP	0.753425	1 / 2t	t=0.664	T	C T
	A20	SNP	0.932203	1 / 2t	t=0.536	T	G T
	A21	SNP	0.859375	1 / 2a	a=0.582	A	G A
	A22	SNP	0.691824	1 / 2g	g=0.723	G	GA
	A23	SNP	1.25143	1 / g	g=0.799	G	G
	A24	SNP	0.873016	1 /2t	t=0.573	T	CT
	A25	SNP	1.19022	1 / a	a=0.84	A	A
	A26	SNP	0.852713	1 / 2t	t=0.586	T	G T
	A27	SNP	1.64662	1 / a	a=0.607	A	A
	A28	SNP	1.19672	1 / g	g=0.836	G	G
	A29	SNP	1.10606	1 / g	g=0.904	G	G
	A30	SNP	1.84034	1 / a	a=0.543	A	A
	A32	SNP	0.866142	1 / 2t	t=0.577	CT	T
	A33	SNP	1.22333	(c+t) / 4ct	c=0.286 t=0.714	C T	C T
	A34	SNP	0.92437	1 / 2g	g=0.541	GC	G
	A36	SNP	1.96429	1 / 2a	a=0.255	AT	A
	A37	SNP	1.47973	1 / t	t=0.676	T	T
	A38	SNP	2.12621	1 / c	c=0.47	C	C
	A39	SNP	0.846154	1 / 2t	t=0.591	A T	T
	A40	SNP	0.743243	1 /2g	g=0.673	GA	G
	A41	SNP	1.19672	1 / c	c=0.836	C	C
	A42	SNP	1.65909	1 / g	g=0.603	G	G
	A43	SNP	1.15263	1 / a	a=0.868	A	A
	A44	SNP	1.4129	1 / c	c=0.708	C	C
	A45	SNP	1.22333	(c+t) / 4ct	c=0.286 t=0.714	CT	CT
	A46	SNP	0.639535	1 / 2g	g=0.782	G	G A
	A48	SNP	0.753425	1 / 2t	t=0.664	T	C T
	A49	SNP	1.24432	1 / g	g=0.804	G	G
	A50	SNP	1.72441	1 / g	g=0.58	G	G
	A51	SNP	1.05769	1 / 2a	a=0.473	G A	A
	A52	SNP	1.3949	1 / a	a=0.717	A	A
	A54	SNP	1.00298	(a+g) / 4ag	a=0.527 g=0.473	G A	GA
	A35	SNP	1.00917	1 / 2a	a=0.495	A	GA
	A53	SNP	1.79508	1 / a	a=0.557	A	A
	cumulative LR		796.423				

Case	3		Scenario	3					
	M		Mother	1000-0102					
	C		Child	1000-0103					
	F		Tested Man	1000-0104					
C: M + F									
/C: M+ ?									
	Malay		cumulative LR	7.55221e6	Posterior probability=	99.99999%	assuming prior=		50%
►	Malay					M	C		F
	A01	SNP	1	1 / (c+t)	c=0.427 t=0.573	C T	C T		T
	A02	SNP	2.82051	1 / t	t=0.355	C T	T		T
	A03	SNP	1.19022	1 / t	t=0.84	T	T		T
	A04	SNP	2.29167	1 / t	t=0.436	A T	T		T
	A05	SNP	3.72881	1 / g	g=0.268	A	G A		G
	A06	SNP	1.51724	1 / t	t=0.659	A T	T		T
	A07	SNP	1	1 / (a+g)	g=0.773 a=0.227	G A	G A		G A
	A08	SNP	1.76	1 / t	t=0.568	C T	T		T
	A09	SNP	0.873016	1 / 2c	c=0.573	C	C		G C
	A10	SNP	1.22222	1 / t	t=0.818	C	C T		T
	A11	SNP	1.92982	1 / 2c	c=0.259	C T	C		C T
	A12	SNP	1	1 / (a+g)	g=0.295 a=0.705	G A	G A		G A
	A13	SNP	1	1 / (a+g)	g=0.855 a=0.145	G A	G A		G A
	A14	SNP	1.14583	1 / 2t	t=0.436	T	T		A T
	A15	SNP	1	1 / (c+t)	c=0.35 t=0.65	C T	C T		C T
	A16	SNP	3.14286	1 / 2t	t=0.159	G	G T		G T
	A17	SNP	1.15263	1 / a	a=0.868	A	A		A
	A18	SNP	1	1 / (g+t)	g=0.6 t=0.4	G T	G T		G T
	A19	SNP	1.48649	1 / 2c	c=0.336	T	C T		C T
	A20	SNP	1	1 / (g+t)	g=0.464 t=0.536	G T	G T		T
	A21	SNP	1	1 / (a+g)	g=0.418 a=0.582	G A	G A		A
	A22	SNP	1.37736	1 / g	g=0.726	G	G		G
	A23	SNP	1	1 / (a+g)	g=0.795 a=0.205	G A	G A		A
	A24	SNP	1.7381	1 / t	t=0.575	T	T		T
	A25	SNP	1	1 / (a+g)	g=0.164 a=0.836	G A	G A		A
	A26	SNP	1.20879	1 / 2g	g=0.414	T	G T		G T
	A27	SNP	1.26437	1 / 2g	g=0.395	A	G A		G A
	A28	SNP	1.19672	1 / g	g=0.836	G	G		G
	A29	SNP	1.10606	1 / g	g=0.904	G	G		G
	A30	SNP	1.84874	1 / a	a=0.541	C	A C		A
	A32	SNP	2.36559	1 / c	c=0.423	T	C T		C
	A33	SNP	1	1 / (c+t)	c=0.286 t=0.714	C T	C T		T
	A34	SNP	1.84874	1 / g	g=0.541	C	G C		G
	A36	SNP	0.670732	1 / 2t	t=0.745	T	T		A T
	A37	SNP	1.48649	1 / t	t=0.673	C	C T		T
	A38	SNP	2.12621	1 / c	c=0.47	C	C		C
	A39	SNP	0.846154	1 / 2t	t=0.591	A T	T		A T
	A40	SNP	1.47973	1 / g	g=0.676	G	G		G
	A41	SNP	0.601093	1 / 2c	c=0.832	C	C		C T
	A42	SNP	2.5	1 / a	a=0.4	G A	A		A
	A43	SNP	3.66667	1 / 2g	g=0.136	G A	G		G A
	A44	SNP	1.4129	1 / c	c=0.708	C	C		C
	A45	SNP	3.47619	1 / t	t=0.288	T	T		T
	A46	SNP	1	1 / (a+g)	g=0.782 a=0.218	G A	G A		A
	A48	SNP	1.50685	1 / t	t=0.664	C	C T		T
	A49	SNP	1.24432	1 / g	g=0.804	G	G		G
	A50	SNP	1	1 / (a+g)	g=0.577 a=0.423	G A	G A		G
	A51	SNP	1	1 / (a+g)	g=0.527 a=0.473	G A	G A		G A
	A52	SNP	0.700637	1 / 2a	a=0.714	A	A		A T
	A54	SNP	0.948276	1 / 2a	a=0.527	G A	A		G A
	A35	SNP	1.97297	1 / g	g=0.507	G	G		G
	A53	SNP	1.12245	1 / 2g	g=0.445	G	G		G A
	cumulative LR		7.55221e6						

Case	4		Scenario	4				
	M		Mother	1000-0105				
	C		Child	1000-0106				
	F		Tested Man	1000-0107				
C: M + F								
/C: M+ ?								
	Malay		cumulative LR	9.45636e6	Posterior probability=	99.99999%	assuming prior=	50%
►	Malay					M	C	F
	A01	SNP	1.74603	1 / t	t=0.573	C T	T	T
	A02	SNP	1.54225	1 / c	c=0.648	C	C	C
	A03	SNP	1.19022	1 / t	t=0.84	T	T	T
	A04	SNP	2.29167	1 / t	t=0.436	A T	T	T
	A05	SNP	3.72881	1 / g	g=0.268	A	G A	G
	A06	SNP	1.46667	1 / 2a	a=0.341	T	A T	A T
	A07	SNP	1.28824	1 / g	g=0.776	G	G	G
	A08	SNP	1	1 / (c+t)	c=0.432 t=0.568	C T	C T	C T
	A09	SNP	2.32979	1 / g	g=0.429	G	G	G
	A10	SNP	1.21667	1 / t	t=0.822	T	T	T
	A11	SNP	1	1 / (c+t)	c=0.259 t=0.741	C T	C T	T
	A12	SNP	3.38462	1 / g	g=0.295	A	G A	G
	A13	SNP	1	1 / (a+g)	g=0.855 a=0.145	G A	G A	G
	A14	SNP	1	1 / (a+t)	a=0.564 t=0.436	A T	A T	T
	A15	SNP	0.769231	1 / 2t	t=0.65	C T	T	C T
	A16	SNP	3.14286	1 / 2t	t=0.159	G	G T	G T
	A17	SNP	1.15263	1 / a	a=0.868	A	A	A
	A18	SNP	1.66667	1 / g	g=0.6	T	G T	G
	A19	SNP	1.5	1 / t	t=0.667	T	T	T
	A20	SNP	1	1 / (g+t)	g=0.464 t=0.536	G T	G T	G T
	A21	SNP	1.71875	1 / a	a=0.582	G A	A	A
	A22	SNP	1	1 / (a+g)	g=0.723 a=0.277	G A	G A	G
	A23	SNP	1.25714	1 / g	g=0.795	G A	G	G
	A24	SNP	0.873016	1 / 2t	t=0.573	C T	T	C T
	A25	SNP	1	1 / (a+g)	g=0.164 a=0.836	G A	G A	A
	A26	SNP	1.20879	1 / 2g	g=0.414	T	G T	G T
	A27	SNP	1	1 / (a+g)	g=0.395 a=0.605	G A	G A	A
	A28	SNP	1.19672	1 / g	g=0.836	G	G	G
	A29	SNP	1.10606	1 / g	g=0.904	G	G	G
	A30	SNP	0.92437	1 / 2a	a=0.541	A	A	A C
	A32	SNP	1.1828	1 / 2c	c=0.423	T	C T	C T
	A33	SNP	0.700637	1 / 2t	t=0.714	C	C T	C T
	A34	SNP	1.84874	1 / g	g=0.541	G C	G	G
	A36	SNP	1	1 / (a+t)	a=0.255 t=0.745	A T	A T	A T
	A37	SNP	1	1 / (c+t)	c=0.327 t=0.673	C T	C T	C T
	A38	SNP	0.940171	1 / 2g	g=0.532	G C	G	G C
	A39	SNP	1	1 / (a+t)	a=0.409 t=0.591	A T	A T	A T
	A40	SNP	1	1 / (a+g)	g=0.673 a=0.327	G A	G A	G A
	A41	SNP	1	1 / (c+t)	c=0.832 t=0.168	C T	C T	C
	A42	SNP	2.5	1 / a	a=0.4	G A	A	A
	A43	SNP	1.15263	1 / a	a=0.868	A	A	A
	A44	SNP	1.4129	1 / c	c=0.708	C	C	C
	A45	SNP	1.74603	1 / 2t	t=0.286	C	C T	C T
	A46	SNP	1	1 / (a+g)	g=0.782 a=0.218	G A	G A	G
	A48	SNP	1.5	1 / t	t=0.667	T	T	T
	A49	SNP	5	1 / a	a=0.2	G	G A	A
	A50	SNP	1.73228	1 / g	g=0.577	A	G A	G
	A51	SNP	1	1 / (a+g)	g=0.527 a=0.473	G A	G A	G A
	A52	SNP	1.74603	1 / 2t	t=0.286	T	T	A T
	A54	SNP	0.948276	1 / 2a	a=0.527	A	A	G A
	A35	SNP	1.98198	1 / g	g=0.505	G A	G	G
	A53	SNP	2.2449	1 / g	g=0.445	A	G A	G
	cumulative LR		9.45636e6					

Case	5		Scenario	5				
	M		Mother	1000-0108				
	C		Child	1000-0109				
	F		Tested Man	1000-0110				
C: M + F								
/C: M+ ?								
	Malay		cumulative LR	14.1494e6	Posterior probability=	99.999993%	assuming prior=	50%

►	Malay					M	C	F
	A01	SNP	1.74603	1 / t	t=0.573	C	C T	T
	A02	SNP	1.54225	1 / c	c=0.648	C	C	C
	A03	SNP	0.597826	1 / 2t	t=0.836	T	T	A T
	A04	SNP	1	1 / (a+t)	a=0.564 t=0.436	A T	A T	A
	A05	SNP	3.72881	1 / g	g=0.268	A	G A	G
	A06	SNP	1.51034	1 / t	t=0.662	T	T	T
	A07	SNP	1	1 / (a+g)	g=0.773 a=0.227	G A	G A	G
	A08	SNP	1.15789	1 / 2c	c=0.432	T	C T	C T
	A09	SNP	1	1 / (c+g)	g=0.427 c=0.573	G C	G C	C
	A10	SNP	1.22222	1 / t	t=0.818	C	C T	T
	A11	SNP	1	1 / (c+t)	c=0.259 t=0.741	C T	C T	C T
	A12	SNP	1.4129	1 / a	a=0.708	A	A	A
	A13	SNP	6.875	1 / a	a=0.145	G A	A	A
	A14	SNP	1.14583	1 / 2t	t=0.436	T	T	A T
	A15	SNP	1.53846	1 / t	t=0.65	C T	T	T
	A16	SNP	3.14286	1 / 2t	t=0.159	G	G T	G T
	A17	SNP	1.15263	1 / a	a=0.868	A	A	A
	A18	SNP	2.5	1 / t	t=0.4	G	G T	T
	A19	SNP	1.48649	1 / 2c	c=0.336	C	C	C T
	A20	SNP	1.07843	1 / 2g	g=0.464	G T	G	G T
	A21	SNP	1.71094	1 / a	a=0.584	A	A	A
	A22	SNP	3.60656	1 / a	a=0.277	G	G A	A
	A23	SNP	1.25143	1 / g	g=0.799	G	G	G
	A24	SNP	1	1 / (c+t)	c=0.427 t=0.573	C T	C T	C T
	A25	SNP	0.597826	1 / 2a	a=0.836	A	A	G A
	A26	SNP	1	1 / (g+t)	g=0.414 t=0.586	G T	G T	G T
	A27	SNP	1	1 / (a+g)	g=0.395 a=0.605	G A	G A	A
	A28	SNP	1.19672	1 / g	g=0.836	G	G	G
	A29	SNP	10	1 / a	a=0.1	G	G A	A
	A30	SNP	1.08911	1 / 2c	c=0.459	C	C	A C
	A32	SNP	1	1 / (c+t)	c=0.423 t=0.577	C T	C T	C T
	A33	SNP	1.74603	1 / 2c	c=0.286	C	C	C T
	A34	SNP	1.08911	1 / 2c	c=0.459	C	C	G C
	A36	SNP	1.96429	1 / 2a	a=0.255	A T	A	A T
	A37	SNP	3.05556	1 / c	c=0.327	C T	C	C
	A38	SNP	2.13592	1 / c	c=0.468	G C	C	C
	A39	SNP	1	1 / (a+t)	a=0.409 t=0.591	A T	A T	A T
	A40	SNP	3.05556	1 / a	a=0.327	G	G A	A
	A41	SNP	0.601093	1 / 2c	c=0.832	C	C	C T
	A42	SNP	1.25	1 / 2a	a=0.4	A	A	G A
	A43	SNP	0.578947	1 / 2a	a=0.864	A	A	G A
	A44	SNP	1	1 / (c+t)	c=0.705 t=0.295	C T	C T	C T
	A45	SNP	1.3949	1 / c	c=0.717	C	C	C
	A46	SNP	1	1 / (a+g)	g=0.782 a=0.218	G A	G A	G
	A48	SNP	1	1 / (c+t)	c=0.336 t=0.664	C T	C T	C
	A49	SNP	0.625	1 / 2g	g=0.8	G	G	G A
	A50	SNP	0.866142	1 / 2g	g=0.577	G	G	G A
	A51	SNP	0.948276	1 / 2g	g=0.527	G	G	G A
	A52	SNP	1.74603	1 / 2t	t=0.286	T	T	A T
	A54	SNP	1	1 / (a+g)	g=0.473 a=0.527	G A	G A	A
	A35	SNP	1.00917	1 / 2a	a=0.495	A	A	G A
	A53	SNP	1.80328	1 / a	a=0.555	G	G A	A
	cumulative LR		14.1494e6					

Case	6		Scenario	6				
	M		Mother	1000-0111				
	C		Child	1000-0112				
	F		Tested Man	1000-0113				
C: M +F								
/C: M+?								
	Malay		cumulative LR	328.849e6	Posterior probability=	99.9999997 %	assuming prior=	50%
►	Malay					M	C	F
	A01	SNP	1.17021	1 / 2c	c=0.427	T	C T	C T
	A02	SNP	1.5493	1 / c	c=0.645	C T	C	C
	A03	SNP	3.05556	1 / 2a	a=0.164	T	A T	A T
	A04	SNP	1.14583	1 / 2t	t=0.436	T	T	A T
	A05	SNP	1.36025	1 / a	a=0.735	A	A	A
	A06	SNP	1.51034	1 / t	t=0.662	T	T	T
	A07	SNP	0.647059	1 / 2g	g=0.773	G	G	G A
	A08	SNP	2.31579	1 / c	c=0.432	T	C T	C
	A09	SNP	2.34043	1 / g	g=0.427	C	G C	G
	A10	SNP	1.21667	1 / t	t=0.822	T	T	T
	A11	SNP	1.92982	1 / 2c	c=0.259	T	C T	C T
	A12	SNP	3.38462	1 / g	g=0.295	A	G A	G
	A13	SNP	1.16489	1 / g	g=0.858	G	G	G
	A14	SNP	1.77419	1 / a	a=0.564	T	A T	A
	A15	SNP	1.53846	1 / t	t=0.65	C	C T	T
	A16	SNP	1.18378	1 / g	g=0.845	G	G	G
	A17	SNP	1.15263	1 / a	a=0.868	A	A	A
	A18	SNP	1.66667	1 / g	g=0.6	T	G T	G
	A19	SNP	1.50685	1 / t	t=0.664	C T	T	T
	A20	SNP	1	1 / (g+t)	g=0.464 t=0.536	G T	G T	G
	A21	SNP	1	1 / (a+g)	g=0.418 a=0.582	G A	G A	A
	A22	SNP	1.37736	1 / g	g=0.726	G	G	G
	A23	SNP	4.88889	1 / a	a=0.205	G A	A	A
	A24	SNP	1.7381	1 / t	t=0.575	T	T	T
	A25	SNP	1.19022	1 / a	a=0.84	A	A	A
	A26	SNP	1	1 / (g+t)	g=0.414 t=0.586	G T	G T	G T
	A27	SNP	1.65414	1 / a	a=0.605	G A	A	A
	A28	SNP	1.19672	1 / g	g=0.836	G	G	G
	A29	SNP	1.11111	1 / g	g=0.9	G A	G	G
	A30	SNP	2.16832	1 / c	c=0.461	C	C	C
	A32	SNP	1	1 / (c+t)	c=0.423 t=0.577	C T	C T	T
	A33	SNP	1.40127	1 / t	t=0.714	C T	T	T
	A34	SNP	1.08911	1 / 2c	c=0.459	G C	C	G C
	A36	SNP	1.34146	1 / t	t=0.745	A T	T	T
	A37	SNP	1.47973	1 / t	t=0.676	T	T	T
	A38	SNP	1.06796	1 / 2c	c=0.468	G	G C	G C
	A39	SNP	1.68462	1 / t	t=0.594	T	T	T
	A40	SNP	1.52778	1 / 2a	a=0.327	G A	A	G A
	A41	SNP	1.19672	1 / c	c=0.836	C	C	C
	A42	SNP	1	1 / (a+g)	g=0.6 a=0.4	G A	G A	G A
	A43	SNP	3.66667	1 / 2g	g=0.136	A	G A	G A
	A44	SNP	1.4129	1 / c	c=0.708	C	C	C
	A45	SNP	1.74603	1 / 2t	t=0.286	C	C T	C T
	A46	SNP	1	1 / (a+g)	g=0.782 a=0.218	G A	G A	G A
	A48	SNP	1.5	1 / t	t=0.667	T	T	T
	A49	SNP	1.24432	1 / g	g=0.804	G	G	G
	A50	SNP	1.1828	1 / 2a	a=0.423	G	G A	G A
	A51	SNP	2.11538	1 / a	a=0.473	G A	A	A
	A52	SNP	0.700637	1 / 2a	a=0.714	A	A	A T
	A54	SNP	1.88793	1 / a	a=0.53	A	A	A
	A35	SNP	1	1 / (a+g)	g=0.505 a=0.495	G A	G A	A
	A53	SNP	2.2449	1 / g	g=0.445	A	G A	G
	cumulative LR		328.849e6					

Case 7 Scenario 7

M Mother 1000-0114

C Child 1000-0115

C: M+?
/C:?+ ?

Malay	cumulative LR	5.1389e6	Posterior probability=	99.99998 %	assuming prior=	50%

► Malay					M	C
A01	SNP	0.873016	1 / 2t	t=0.573	C T	T
A02	SNP	1.09245	(c+t) / 4ct	c=0.645 t=0.355	C T	C T
A03	SNP	6.08333	1 / a	a=0.164	A	A
A04	SNP	0.887097	1 / 2a	a=0.564	A	A T
A05	SNP	0.68323	1 / 2a	a=0.732	G A	A
A06	SNP	0.758621	1 / 2t	t=0.659	A T	T
A07	SNP	1.28824	1 / g	g=0.776	G	G
A08	SNP	1.752	1 / t	t=0.571	T	T
A09	SNP	1.02161	(c+g) / 4cg	g=0.427 c=0.573	G C	G C
A10	SNP	0.611111	1 / 2t	t=0.818	C T	T
A11	SNP	1.92982	1 / 2c	c=0.259	C T	C
A12	SNP	1.20099	(a+g) / 4ag	g=0.295 a=0.705	G A	G A
A13	SNP	1.16489	1 / g	g=0.858	G	G
A14	SNP	0.887097	1 / 2a	a=0.564	A T	A
A15	SNP	1.53147	1 / t	t=0.653	T	T
A16	SNP	3.14286	1 / 2t	t=0.159	T	G T
A17	SNP	1.15263	1 / a	a=0.868	A	A
A18	SNP	1.65909	1 / g	g=0.603	G	G
A19	SNP	1.11996	(c+t) / 4ct	c=0.336 t=0.664	C T	C T
A20	SNP	1.07843	1 / 2g	g=0.464	G	G T
A21	SNP	0.859375	1 / 2a	a=0.582	G A	A
A22	SNP	1.37736	1 / g	g=0.726	G	G
A23	SNP	2.44444	1 / 2a	a=0.205	G A	A
A24	SNP	0.873016	1 / 2t	t=0.573	T	C T
A25	SNP	1.82669	(a+g) / 4ag	g=0.164 a=0.836	G A	G A
A26	SNP	1.03075	(g+t) / 4gt	g=0.414 t=0.586	G T	G T
A27	SNP	1.64662	1 / a	a=0.607	A	A
A28	SNP	1.19672	1 / g	g=0.836	G	G
A29	SNP	5	1 / 2a	a=0.1	G A	A
A30	SNP	1.84034	1 / a	a=0.543	A	A
A32	SNP	1.1828	1 / 2c	c=0.423	C T	C
A33	SNP	1.3949	1 / t	t=0.717	T	T
A34	SNP	1.00674	(c+g) / 4cg	g=0.541 c=0.459	G C	G C
A36	SNP	0.670732	1 / 2t	t=0.745	T	A T
A37	SNP	1.13551	(c+t) / 4ct	c=0.327 t=0.673	C T	C T
A38	SNP	1.87179	1 / g	g=0.534	G	G
A39	SNP	1.03419	(a+t) / 4at	a=0.409 t=0.591	A T	A T
A40	SNP	1.47973	1 / g	g=0.676	G	G
A41	SNP	2.97297	1 / 2t	t=0.168	T	C T
A42	SNP	0.833333	1 / 2g	g=0.6	G	G A
A43	SNP	0.578947	1 / 2a	a=0.864	G A	A
A44	SNP	1.20099	(c+t) / 4ct	c=0.705 t=0.295	C T	C T
A45	SNP	1.3949	1 / c	c=0.717	C	C
A46	SNP	2.29167	1 / 2a	a=0.218	G A	A
A48	SNP	2.95946	1 / c	c=0.338	C	C
A49	SNP	1.24432	1 / g	g=0.804	G	G
A50	SNP	1.72441	1 / g	g=0.58	G	G
A51	SNP	1.88793	1 / g	g=0.53	G	G
A52	SNP	1.3949	1 / a	a=0.717	A	A
A54	SNP	1.00298	(a+g) / 4ag	g=0.473 a=0.527	G A	G A
A35	SNP	0.990991	1 / 2g	g=0.505	G A	G
A53	SNP	1.79508	1 / a	a=0.557	A	A
cumulative LR		5.1389e6				

Case	8C		Scenario	Control1
	M		Mother	1000-0116
	C		Child #1	1000-0117
	D		Child #2	1000-0118
	E		Child #3	1000-0119
	F		Tested Man	1000-0120

C : M +F
/C: M+?

Malay	cumulative LR	3.96072e6	Posterior probability=	99.99997%	assuming prior=	50%

► Malay					M	C	D	E	F
A01	SNP	2.32979	1 / c	c=0.429	C	C	C	C	C
A02	SNP	1.5493	1 / c	c=0.645	C T	C	C T	C	C
A03	SNP	6.11111	1 / a	a=0.164	T	A T	A T	A T	A
A04	SNP	1.14583	1 / 2t	t=0.436	T	T	T	A T	A T
A05	SNP	0.68323	1 / 2a	a=0.732	G A	A	G	G A	G A
A06	SNP	1.51724	1 / t	t=0.659	A T	T	A T	T	T
A07	SNP	1	1 / (a+g)	g=0.773 a=0.227	G A	G A	G A	G A	G A
A08	SNP	2.31579	1 / c	c=0.432	T	C T	C T	C T	C
A09	SNP	2.32979	1 / g	g=0.429	G	G	G	G	G
A10	SNP	1.22222	1 / t	t=0.818	C T	T	T	T	T
A11	SNP	1	1 / (c+t)	c=0.259 t=0.741	C T	C T	C T	C	C
A12	SNP	1.4129	1 / a	a=0.708	A	A	A	A	A
A13	SNP	1.16489	1 / g	g=0.858	G	G	G	G	G
A14	SNP	1	1 / (a+t)	a=0.564 t=0.436	A T	A T	T	T	A T
A15	SNP	1.53147	1 / t	t=0.653	T	T	T	T	T
A16	SNP	1.18378	1 / g	g=0.845	G	G	G	G	G
A17	SNP	1	1 / (a+g)	g=0.136 a=0.864	G A	G A	G	A	G A
A18	SNP	1	1 / (g+t)	g=0.6 t=0.4	G T	G T	G	G T	G
A19	SNP	1.5	1 / t	t=0.667	T	T	T	T	T
A20	SNP	1.85593	1 / t	t=0.539	T	T	T	T	T
A21	SNP	1.19565	1 / 2g	g=0.418	G	G	G	G A	G A
A22	SNP	0.691824	1 / 2g	g=0.723	G	G	G A	G	G A
A23	SNP	1.25714	1 / g	g=0.795	G A	G	G	G	G
A24	SNP	1.17021	1 / 2c	c=0.427	C T	C	C	T	C T
A25	SNP	3.05556	1 / 2g	g=0.164	A	G A	G A	G A	G A
A26	SNP	0.852713	1 / 2t	t=0.586	T	T	T	T	G T
A27	SNP	1.26437	1 / 2g	g=0.395	A	G A	G A	G A	G A
A28	SNP	1.19672	1 / g	g=0.836	G	G	G	G	G
A29	SNP	1.10606	1 / g	g=0.904	G	G	G	G	G
A30	SNP	1.84034	1 / a	a=0.543	A	A	A	A	A
A32	SNP	1	1 / (c+t)	c=0.423 t=0.577	C T	C T	C T	C T	C T
A33	SNP	3.49206	1 / c	c=0.286	C T	C	C T	C	C
A34	SNP	0.92437	1 / 2g	g=0.541	G	G	G	G	G C
A36	SNP	1.96429	1 / 2a	a=0.255	A T	A	A T	A T	A T
A37	SNP	1.48649	1 / t	t=0.673	C T	T	T	C T	T
A38	SNP	0.940171	1 / 2g	g=0.532	C	G C	C	G	G C
A39	SNP	1.22222	1 / 2a	a=0.409	A	A	A	A T	A T
A40	SNP	0.743243	1 / 2g	g=0.673	G A	G	G A	G A	G A
A41	SNP	0.601093	1 / 2c	c=0.832	C	C	C T	C T	C T
A42	SNP	2.5	1 / a	a=0.4	G A	A	A	G A	A
A43	SNP	0.578947	1 / 2a	a=0.864	A	A	G A	A	G A
A44	SNP	1	1 / (c+t)	c=0.705 t=0.295	C T	C T	C T	C	C
A45	SNP	1.74603	1 / 2t	t=0.286	T	T	C T	T	C T
A46	SNP	1.27326	1 / g	g=0.785	G	G	G	G	G
A48	SNP	1.50685	1 / t	t=0.664	C	C T	C T	C T	T
A49	SNP	1	1 / (a+g)	g=0.8 a=0.2	G A	G A	A	G A	G A
A50	SNP	1.1828	1 / 2a	a=0.423	G	G A	G A	G A	G A
A51	SNP	2.11538	1 / a	a=0.473	G	G A	G A	G A	A
A52	SNP	1.3949	1 / a	a=0.717	A	A	A	A	A
A54	SNP	1.89655	1 / a	a=0.527	G	G A	G A	G A	A
A35	SNP	1.00917	1 / 2a	a=0.495	A	A	A	A	G A
A53	SNP	1.80328	1 / a	a=0.555	G A	A	G A	G A	A
cumulative LR		3.96072e6							

Case	8D		Scenario	Control1						
	M		Mother	1000-0116						
	C		Child #1	1000-0117						
	D		Child #2	1000-0118						
	E		Child #3	1000-0119						
	F		Tested Man	1000-0120						
D:M + F										
/D:M+ ?										
	Malay		cumulative LR	325.836e6	Posterior probability=	99.9999997%	assuming prior=	50%		
►	Malay					M	C	D	E	F
	A01	SNP	2.32979	1 / c	c=0.429	C	C	C	C	C
	A02	SNP	1	1 / (c+t)	c=0.645 t=0.355	C T	C	C T	C	C
	A03	SNP	6.11111	1 / a	a=0.164	T	A T	A T	A T	A
	A04	SNP	1.14583	1 / 2t	t=0.436	T	T	T	A T	A T
	A05	SNP	1.86441	1 / 2g	g=0.268	G A	A	G	G A	G A
	A06	SNP	1	1 / (a+t)	a=0.341 t=0.659	A T	T	A T	T	T
	A07	SNP	1	1 / (a+g)	g=0.773 a=0.227	G A	G A	G A	G A	G A
	A08	SNP	2.31579	1 / c	c=0.432	T	C T	C T	C T	C
	A09	SNP	2.32979	1 / g	g=0.429	G	G	G	G	G
	A10	SNP	1.22222	1 / t	t=0.818	C T	T	T	T	T
	A11	SNP	1	1 / (c+t)	c=0.259 t=0.741	C T	C T	C T	C	C
	A12	SNP	1.4129	1 / a	a=0.708	A	A	A	A	A
	A13	SNP	1.16489	1 / g	g=0.858	G	G	G	G	G
	A14	SNP	1.14583	1 / 2t	t=0.436	A T	A T	T	T	A T
	A15	SNP	1.53147	1 / t	t=0.653	T	T	T	T	T
	A16	SNP	1.18378	1 / g	g=0.845	G	G	G	G	G
	A17	SNP	3.66667	1 / 2g	g=0.136	G A	G A	G	A	G A
	A18	SNP	1.66667	1 / g	g=0.6	G T	G T	G	G T	G
	A19	SNP	1.5	1 / t	t=0.667	T	T	T	T	T
	A20	SNP	1.85593	1 / t	t=0.539	T	T	T	T	T
	A21	SNP	1.19565	1 / 2g	g=0.418	G	G	G	G A	G A
	A22	SNP	1.80328	1 / 2a	a=0.277	G	G	G A	G	G A
	A23	SNP	1.25714	1 / g	g=0.795	G A	G	G	G	G
	A24	SNP	1.17021	1 / 2c	c=0.427	C T	C	C	T	C T
	A25	SNP	3.05556	1 / 2g	g=0.164	A	G A	G A	G A	G A
	A26	SNP	0.852713	1 / 2t	t=0.586	T	T	T	T	G T
	A27	SNP	1.26437	1 / 2g	g=0.395	A	G A	G A	G A	G A
	A28	SNP	1.19672	1 / g	g=0.836	G	G	G	G	G
	A29	SNP	1.10606	1 / g	g=0.904	G	G	G	G	G
	A30	SNP	1.84034	1 / a	a=0.543	A	A	A	A	A
	A32	SNP	1	1 / (c+t)	c=0.423 t=0.577	C T	C T	C T	C T	C T
	A33	SNP	1	1 / (c+t)	c=0.286 t=0.714	C T	C	C T	C	C
	A34	SNP	0.92437	1 / 2g	g=0.541	G	G	G	G	G C
	A36	SNP	1	1 / (a+t)	a=0.255 t=0.745	A T	A	A T	A T	A T
	A37	SNP	1.48649	1 / t	t=0.673	C T	T	T	C T	T
	A38	SNP	1.06796	1 / 2c	c=0.468	C	G C	C	G	G C
	A39	SNP	1.22222	1 / 2a	a=0.409	A	A	A	A T	A T
	A40	SNP	1	1 / (a+g)	g=0.673 a=0.327	G A	G	G A	G A	G A
	A41	SNP	2.97297	1 / 2t	t=0.168	C	C	C T	C T	C T
	A42	SNP	2.5	1 / a	a=0.4	G A	A	A	G A	A
	A43	SNP	3.66667	1 / 2g	g=0.136	A	A	G A	A	G A
	A44	SNP	1	1 / (c+t)	c=0.705 t=0.295	C T	C T	C T	C	C
	A45	SNP	0.700637	1 / 2c	c=0.714	T	T	C T	T	C T
	A46	SNP	1.27326	1 / g	g=0.785	G	G	G	G	G
	A48	SNP	1.50685	1 / t	t=0.664	C	C T	C T	C T	T
	A49	SNP	2.5	1 / 2a	a=0.2	G A	G A	A	G A	G A
	A50	SNP	1.1828	1 / 2a	a=0.423	G	G A	G A	G A	G A
	A51	SNP	2.11538	1 / a	a=0.473	G	G A	G A	G A	A
	A52	SNP	1.3949	1 / a	a=0.717	A	A	A	A	A
	A54	SNP	1.89655	1 / a	a=0.527	G	G A	G A	G A	A
	A35	SNP	1.00917	1 / 2a	a=0.495	A	A	A	A	G A
	A53	SNP	1	1 / (a+g)	g=0.445 a=0.555	G A	A	G A	G A	A
	cumulative LR		325.836e6							

Case	8E	Scenario	Control1
	M	Mother	1000-0116
	C	Child #1	1000-0117
	D	Child #2	1000-0118
	E	Child #3	1000-0119
	F	Tested Man	1000-0120

E: M + F
/E:M+ ?

Malay cumulative LR 3.25562e6 Posterior probability= 99.99997% assuming prior= 50%

► Malay					M	C	D	E	F
A01	SNP	2.32979	1 / c	c=0.429	C	C	C	C	C
A02	SNP	1.5493	1 / c	c=0.645	C T	C	C T	C	C
A03	SNP	6.11111	1 / a	a=0.164	T	A T	A T	A T	A
A04	SNP	0.887097	1 / 2a	a=0.564	T	T	T	A T	A T
A05	SNP	1	1 / (a+g)	g=0.268 a=0.732	G A	A	G	G A	G A
A06	SNP	1.51724	1 / t	t=0.659	A T	T	A T	T	T
A07	SNP	1	1 / (a+g)	g=0.773 a=0.227	G A	G A	G A	G A	G A
A08	SNP	2.31579	1 / c	c=0.432	T	C T	C T	C T	C
A09	SNP	2.32979	1 / g	g=0.429	G	G	G	G	G
A10	SNP	1.22222	1 / t	t=0.818	C T	T	T	T	T
A11	SNP	3.85965	1 / c	c=0.259	C T	C T	C T	C	C
A12	SNP	1.4129	1 / a	a=0.708	A	A	A	A	A
A13	SNP	1.16489	1 / g	g=0.858	G	G	G	G	G
A14	SNP	1.14583	1 / 2t	t=0.436	A T	A T	T	T	A T
A15	SNP	1.53147	1 / t	t=0.653	T	T	T	T	T
A16	SNP	1.18378	1 / g	g=0.845	G	G	G	G	G
A17	SNP	0.578947	1 / 2a	a=0.864	G A	G A	G	A	G A
A18	SNP	1	1 / (g+t)	g=0.6 t=0.4	G T	G T	G	G T	G
A19	SNP	1.5	1 / t	t=0.667	T	T	T	T	T
A20	SNP	1.85593	1 / t	t=0.539	T	T	T	T	T
A21	SNP	0.859375	1 / 2a	a=0.582	G	G	G	G A	G A
A22	SNP	0.691824	1 / 2g	g=0.723	G	G	G A	G	G A
A23	SNP	1.25714	1 / g	g=0.795	G A	G	G	G	G
A24	SNP	0.873016	1 / 2t	t=0.573	C T	C	C	T	C T
A25	SNP	3.05556	1 / 2g	g=0.164	A	G A	G A	G A	G A
A26	SNP	0.852713	1 / 2t	t=0.586	T	T	T	T	G T
A27	SNP	1.26437	1 / 2g	g=0.395	A	G A	G A	G A	G A
A28	SNP	1.19672	1 / g	g=0.836	G	G	G	G	G
A29	SNP	1.10606	1 / g	g=0.904	G	G	G	G	G
A30	SNP	1.84034	1 / a	a=0.543	A	A	A	A	A
A32	SNP	1	1 / (c+t)	c=0.423 t=0.577	C T	C T	C T	C T	C T
A33	SNP	3.49206	1 / c	c=0.286	C T	C	C T	C	C
A34	SNP	0.92437	1 / 2g	g=0.541	G	G	G	G	G C
A36	SNP	1	1 / (a+t)	a=0.255 t=0.745	A T	A	A T	A T	A T
A37	SNP	1	1 / (c+t)	c=0.327 t=0.673	C T	T	T	C T	T
A38	SNP	undefined	0 / 0		C	G C	C	G	G C
A39	SNP	0.846154	1 / 2t	t=0.591	A	A	A	A T	A T
A40	SNP	1	1 / (a+g)	g=0.673 a=0.327	G A	G	G A	G A	G A
A41	SNP	2.97297	1 / 2t	t=0.168	C	C	C T	C T	C T
A42	SNP	1	1 / (a+g)	g=0.6 a=0.4	G A	A	A	G A	A
A43	SNP	0.578947	1 / 2a	a=0.864	A	A	G A	A	G A
A44	SNP	1.41935	1 / c	c=0.705	C T	C T	C T	C	C
A45	SNP	1.74603	1 / 2t	t=0.286	T	T	C T	T	C T
A46	SNP	1.27326	1 / g	g=0.785	G	G	G	G	G
A48	SNP	1.50685	1 / t	t=0.664	C	C T	C T	C T	T
A49	SNP	1	1 / (a+g)	g=0.8 a=0.2	G A	G A	A	G A	G A
A50	SNP	1.1828	1 / 2a	a=0.423	G	G A	G A	G A	G A
A51	SNP	2.11538	1 / a	a=0.473	G	G A	G A	G A	A
A52	SNP	1.3949	1 / a	a=0.717	A	A	A	A	A
A54	SNP	1.89655	1 / a	a=0.527	G	G A	G A	G A	A
A35	SNP	1.00917	1 / 2a	a=0.495	A	A	A	A	G A
A53	SNP	1	1 / (a+g)	g=0.445 a=0.555	G A	A	G A	G A	A
cumulative LR		3.25562e6							

Case	9		Scenario	Control 2	C : M + ?			
	M		Mother	1000-0121	/C: G + ?			
	C		Child	1000-0122				
	G		Man #2	1000-0123				
	Malay		LR=inf.inf.....	1.26564	Posterior probability=	100%	assuming prior=	50%
►	Malay					M	C	G
	A01	SNP	1	1 / 1		T	T	T
	A02	SNP	0.709091	2t / (c+t)	c=0.645 t=0.355	C	C T	C T
	A03	SNP	1	1 / 1		A T	A T	A T
	A04	SNP	1	1 / 1		T	T	T
	A05	SNP	1	1 / 1		A	G A	A
	A06	SNP	0.758621	(a+t) / 2t	a=0.341 t=0.659	A T	A T	A
	A07	SNP	1	1 / 1		G	G A	G
	A08	SNP	infinite	1 / 0		T	T	C
	A09	SNP	1	1 / 1		G C	G C	G C
	A10	SNP	1	1 / 1		T	T	T
	A11	SNP	1	1 / 1		T	T	T
	A12	SNP	1	1 / 1		A	A	A
	A13	SNP	1	1 / 1		G	G	G
	A14	SNP	0.887097	(a+t) / 2a	a=0.564 t=0.436	A T	A T	T
	A15	SNP	infinite	1 / 0		C T	C	T
	A16	SNP	1	1 / 1		G T	T	G T
	A17	SNP	1	1 / 1		G A	A	G A
	A18	SNP	1.2	2g / (g+t)	g=0.6 t=0.4	T	G T	G T
	A19	SNP	1	1 / 1		T	T	T
	A20	SNP	1.07273	2t / (g+t)	g=0.464 t=0.536	G	G T	G T
	A21	SNP	0.859375	(a+g) / 2a	g=0.418 a=0.582	G A	G A	G
	A22	SNP	1	1 / 1		G	G A	G
	A23	SNP	1	1 / 1		G	G	G
	A24	SNP	1	1 / 1		C	C	C
	A25	SNP	infinite	1 / 0		G A	A	G
	A26	SNP	0.852713	(g+t) / 2t	g=0.414 t=0.586	G T	G T	G
	A27	SNP	infinite	1 / 0		A	A	G
	A28	SNP	1	1 / 1		G	G	G
	A29	SNP	1	1 / 1		G	G	G
	A30	SNP	1	1 / 1		A C	A C	A C
	A32	SNP	1.15455	2t / (c+t)	c=0.423 t=0.577	C	C T	C T
	A33	SNP	1	1 / 1		C T	T	C T
	A34	SNP	0.92437	(c+g) / 2g	g=0.541 c=0.459	G C	G C	C
	A36	SNP	1	1 / 1		T	A T	T
	A37	SNP	1	1 / 1		C T	T	C T
	A38	SNP	0.5	1 / 2		G C	C	C
	A39	SNP	infinite	1 / 0		A	A	T
	A40	SNP	1	1 / 1		G A	G A	G A
	A41	SNP	1	1 / 1		C	C T	C
	A42	SNP	infinite	1 / 0		G A	A	G
	A43	SNP	2	2 / 1		A	A	G A
	A44	SNP	2	2 / 1		C	C	C T
	A45	SNP	1	1 / 1		C	C	C
	A46	SNP	2.29167	(a+g) / 2a	g=0.782 a=0.218	G A	G A	G
	A48	SNP	1.32727	2t / (c+t)	c=0.336 t=0.664	C	C T	C T
	A49	SNP	1	1 / 1		G	G	G
	A50	SNP	0.866142	(a+g) / 2g	g=0.577 a=0.423	G A	G A	A
	A51	SNP	infinite	1 / 0		G A	G	A
	A52	SNP	1	1 / 1		A	A T	A
	A54	SNP	1	1 / 1		G A	G	G A
	A35	SNP	0.5	1 / 2		G A	G	G
	A53	SNP	1	1 / 1		G A	G A	G A
	LR=inf.inf.....		1.26564					

Case	10	Scenario	10				
	C	Child	1000-00124				
	F	Tested Man	1000-00125				
C: F + ?							
/C: ? + ?							
	Malay	cumulative LR	77.3155	Posterior probability=	98.7%	assuming prior=	50%

►	Malay					C	F
	A01	SNP	1.17021	1 / 2c	c=0.427	C T	C
	A02	SNP	1.54225	1 / c	c=0.648	C	C
	A03	SNP	1.19022	1 / t	t=0.84	T	T
	A04	SNP	2.28125	1 / t	t=0.438	T	T
	A05	SNP	0.68323	1 / 2a	a=0.732	A	G A
	A06	SNP	1.51034	1 / t	t=0.662	T	T
	A07	SNP	1.42353	(a+g) / 4ag	g=0.773 a=0.227	G A	G A
	A08	SNP	1.15789	1 / 2c	c=0.432	C T	C
	A09	SNP	1.7381	1 / c	c=0.575	C	C
	A10	SNP	0.611111	1 / 2t	t=0.818	T	C T
	A11	SNP	0.674847	1 / 2t	t=0.741	T	C T
	A12	SNP	1.20099	(a+g) / 4ag	g=0.295 a=0.705	G A	G A
	A13	SNP	0.585106	1 / 2g	g=0.855	G	G A
	A14	SNP	1.01647	(a+t) / 4at	a=0.564 t=0.436	A T	A T
	A15	SNP	1.0989	(c+t) / 4ct	c=0.35 t=0.65	C T	C T
	A16	SNP	0.594595	1 / 2g	g=0.841	G	G T
	A17	SNP	1.15263	1 / a	a=0.868	A	A
	A18	SNP	1.25	1 / 2t	t=0.4	T	G T
	A19	SNP	1.5	1 / t	t=0.667	T	T
	A20	SNP	1.00532	(g+t) / 4gt	g=0.464 t=0.536	G T	G T
	A21	SNP	1.19565	1 / 2g	g=0.418	G A	G
	A22	SNP	0.691824	1 / 2g	g=0.723	G A	G
	A23	SNP	1.25143	1 / g	g=0.799	G	G
	A24	SNP	1.02161	(c+t) / 4ct	c=0.427 t=0.573	C T	C T
	A25	SNP	1.19022	1 / a	a=0.84	A	A
	A26	SNP	1.03075	(g+t) / 4gt	g=0.414 t=0.586	G T	G T
	A27	SNP	1.04572	(a+g) / 4ag	g=0.395 a=0.605	G A	G A
	A28	SNP	1.78703	(a+g) / 4ag	g=0.832 a=0.168	G A	G A
	A29	SNP	1.10606	1 / g	g=0.904	G	G
	A30	SNP	1.08911	1 / 2c	c=0.459	C	A C
	A32	SNP	1.02447	(c+t) / 4ct	c=0.423 t=0.577	C T	C T
	A33	SNP	1.22333	(c+t) / 4ct	c=0.286 t=0.714	C T	C T
	A34	SNP	1.08911	1 / 2c	c=0.459	C	G C
	A36	SNP	1.31751	(a+t) / 4at	a=0.255 t=0.745	A T	A T
	A37	SNP	0.743243	1 / 2t	t=0.673	C T	T
	A38	SNP	0.940171	1 / 2g	g=0.532	G C	G
	A39	SNP	1.03419	(a+t) / 4at	a=0.409 t=0.591	A T	A T
	A40	SNP	1.13551	(a+g) / 4ag	g=0.673 a=0.327	G A	G A
	A41	SNP	1.19672	1 / c	c=0.836	C	C
	A42	SNP	1.04167	(a+g) / 4ag	g=0.6 a=0.4	G A	G A
	A43	SNP	0.578947	1 / 2a	a=0.864	A	G A
	A44	SNP	1.4129	1 / c	c=0.708	C	C
	A45	SNP	1.3949	1 / c	c=0.717	C	C
	A46	SNP	0.639535	1 / 2g	g=0.782	G	G A
	A48	SNP	1.11996	(c+t) / 4ct	c=0.336 t=0.664	C T	C T
	A49	SNP	1.24432	1 / g	g=0.804	G	G
	A50	SNP	2.35484	1 / a	a=0.425	A	A
	A51	SNP	0.948276	1 / 2g	g=0.527	G A	G
	A52	SNP	1.22333	(a+t) / 4at	a=0.714 t=0.286	A T	A T
	A54	SNP	0.948276	1 / 2a	a=0.527	G A	A
	A35	SNP	1.00008	(a+g) / 4ag	g=0.505 a=0.495	G A	G A
	A53	SNP	0.901639	1 / 2a	a=0.555	G A	A
	cumulative LR		77.3155				

Case	11		Scenario	11					
	M		Mother	1000-0126					
	C		Child	1000-0127					
C: M + ?									
/C: ? + ?									
	Malay		cumulative LR	409.145	Posterior probability=	99.8%	assuming prior=	50%
►	Malay					M	C	
	A01	SNP	1.7381	1 / t	t=0.575	T	T	
	A02	SNP	1.54225	1 / c	c=0.648	C	C	
	A03	SNP	1.19022	1 / t	t=0.84	T	T	
	A04	SNP	1.14583	1 / 2t	t=0.436	A T	T	
	A05	SNP	1.36025	1 / a	a=0.735	A	A	
	A06	SNP	0.758621	1 / 2t	t=0.659	A T	T	
	A07	SNP	1.28824	1 / g	g=0.776	G	G	
	A08	SNP	1.15789	1 / 2c	c=0.432	C	C T	
	A09	SNP	0.873016	1 / 2c	c=0.573	G C	C	
	A10	SNP	1.21667	1 / t	t=0.822	T	T	
	A11	SNP	0.674847	1 / 2t	t=0.741	T	C T	
	A12	SNP	1.69231	1 / 2g	g=0.295	G A	G	
	A13	SNP	1.16489	1 / g	g=0.858	G	G	
	A14	SNP	0.887097	1 / 2a	a=0.564	A	A T	
	A15	SNP	0.769231	1 / 2t	t=0.65	C T	T	
	A16	SNP	1.18378	1 / g	g=0.845	G	G	
	A17	SNP	0.578947	1 / 2a	a=0.864	A	G A	
	A18	SNP	1.25	1 / 2t	t=0.4	T	G T	
	A19	SNP	1.11996	(c+t) / 4ct	c=0.336 t=0.664	C T	C T	
	A20	SNP	1.07843	1 / 2g	g=0.464	G	G T	
	A21	SNP	0.859375	1 / 2a	a=0.582	G A	A	
	A22	SNP	0.691824	1 / 2g	g=0.723	G A	G	
	A23	SNP	0.628571	1 / 2g	g=0.795	G	G A	
	A24	SNP	1.17021	1 / 2c	c=0.427	C	C T	
	A25	SNP	1.19022	1 / a	a=0.84	A	A	
	A26	SNP	1.03075	(g+t) / 4gt	g=0.414 t=0.586	G T	G T	
	A27	SNP	1.26437	1 / 2g	g=0.395	G A	G	
	A28	SNP	1.19672	1 / g	g=0.836	G	G	
	A29	SNP	0.555556	1 / 2g	g=0.9	G	G A	
	A30	SNP	1.08911	1 / 2c	c=0.459	A C	C	
	A32	SNP	1.72441	1 / t	t=0.58	T	T	
	A33	SNP	1.3949	1 / t	t=0.717	T	T	
	A34	SNP	0.92437	1 / 2g	g=0.541	G	G C	
	A36	SNP	1.31751	(a+t) / 4at	a=0.255 t=0.745	A T	A T	
	A37	SNP	0.743243	1 / 2t	t=0.673	C T	T	
	A38	SNP	0.940171	1 / 2g	g=0.532	G C	G	
	A39	SNP	1.03419	(a+t) / 4at	a=0.409 t=0.591	A T	A T	
	A40	SNP	1.52778	1 / 2a	a=0.327	A	G A	
	A41	SNP	1.19672	1 / c	c=0.836	C	C	
	A42	SNP	1.25	1 / 2a	a=0.4	G A	A	
	A43	SNP	2.12281	(a+g) / 4ag	g=0.136 a=0.864	G A	G A	
	A44	SNP	1.4129	1 / c	c=0.708	C	C	
	A45	SNP	1.3949	1 / c	c=0.717	C	C	
	A46	SNP	1.27326	1 / g	g=0.785	G	G	
	A48	SNP	0.753425	1 / 2t	t=0.664	C T	T	
	A49	SNP	1.24432	1 / g	g=0.804	G	G	
	A50	SNP	1.72441	1 / g	g=0.58	G	G	
	A51	SNP	0.948276	1 / 2g	g=0.527	G A	G	
	A52	SNP	1.74603	1 / 2t	t=0.286	A T	T	
	A54	SNP	1.05769	1 / 2g	g=0.473	G	G A	
	A35	SNP	2.00917	1 / a	a=0.498	A	A	
	A53	SNP	1.01204	(a+g) / 4ag	g=0.445 a=0.555	G A	G A	
	cumulative LR		409.145					

Case	12		Scenario	12				
	M		Mother	1000-0220				
	C		Child	1000-0221				
	F		Tested Man	1000-0222				
C: M + F /C:M+ ?								
	Malay		cumulative LR	17.426e6	Posterior probability=	99.999994 %	assuming prior=	50 %
►	Malay					M	C	F
	A01	SNP	0.873016	1 / 2t	t=0.573	T	T	C T
	A02	SNP	1	1 / (c+t)	c=0.645 t=0.355	C T	C T	T
	A03	SNP	1.19022	1 / t	t=0.84	T	T	T
	A04	SNP	2.29167	1 / t	t=0.436	A	A T	T
	A05	SNP	1.36025	1 / a	a=0.735	A	A	A
	A06	SNP	1.51724	1 / t	t=0.659	A	A T	T
	A07	SNP	1.28824	1 / g	g=0.776	G	G	G
	A08	SNP	1	1 / (c+t)	c=0.432 t=0.568	C T	C T	T
	A09	SNP	0.873016	1 / 2c	c=0.573	C	C	G C
	A10	SNP	2.75	1 / 2c	c=0.182	T	C T	C T
	A11	SNP	1.34356	1 / t	t=0.744	T	T	T
	A12	SNP	0.709677	1 / 2a	a=0.705	G A	A	G A
	A13	SNP	1.17021	1 / g	g=0.855	G A	G	G
	A14	SNP	1.77419	1 / a	a=0.564	A T	A	A
	A15	SNP	1.53147	1 / t	t=0.653	T	T	T
	A16	SNP	1.18378	1 / g	g=0.845	G	G	G
	A17	SNP	1.15263	1 / a	a=0.868	A	A	A
	A18	SNP	1.25	1 / 2t	t=0.4	T	T	G T
	A19	SNP	1.48649	1 / 2c	c=0.336	C T	C	C T
	A20	SNP	1.07843	1 / 2g	g=0.464	G	G	G T
	A21	SNP	2.3913	1 / g	g=0.418	A	G A	G
	A22	SNP	3.60656	1 / a	a=0.277	G	G A	A
	A23	SNP	2.44444	1 / 2a	a=0.205	G	G A	G A
	A24	SNP	2.34043	1 / c	c=0.427	T	C T	C
	A25	SNP	1.19022	1 / a	a=0.84	A	A	A
	A26	SNP	1	1 / (g+t)	g=0.414 t=0.586	G T	G T	G T
	A27	SNP	1.26437	1 / 2g	g=0.395	G A	G	G A
	A28	SNP	1.19672	1 / g	g=0.836	G	G	G
	A29	SNP	0.555556	1 / 2g	g=0.9	G A	G	G A
	A30	SNP	1.84034	1 / a	a=0.543	A	A	A
	A32	SNP	1.72441	1 / t	t=0.58	T	T	T
	A33	SNP	1	1 / (c+t)	c=0.286 t=0.714	C T	C T	C T
	A34	SNP	1.84034	1 / g	g=0.543	G	G	G
	A36	SNP	1.96429	1 / 2a	a=0.255	T	A T	A T
	A37	SNP	1.52778	1 / 2c	c=0.327	T	C T	C T
	A38	SNP	2.12621	1 / c	c=0.47	C	C	C
	A39	SNP	1.22222	1 / 2a	a=0.409	A	A	A T
	A40	SNP	1	1 / (a+g)	g=0.673 a=0.327	G A	G A	G
	A41	SNP	1.19672	1 / c	c=0.836	C	C	C
	A42	SNP	0.833333	1 / 2g	g=0.6	A	G A	G A
	A43	SNP	3.66667	1 / 2g	g=0.136	A	G A	G A
	A44	SNP	1	1 / (c+t)	c=0.705 t=0.295	C T	C T	T
	A45	SNP	1.3949	1 / c	c=0.717	C	C	C
	A46	SNP	1.27326	1 / g	g=0.785	G	G	G
	A48	SNP	1.48649	1 / 2c	c=0.336	T	C T	C T
	A49	SNP	1.24432	1 / g	g=0.804	G	G	G
	A50	SNP	1.73228	1 / g	g=0.577	G A	G	G
	A51	SNP	0.948276	1 / 2g	g=0.527	G A	G	G A
	A52	SNP	1.40127	1 / a	a=0.714	A T	A	A
	A54	SNP	1	1 / (a+g)	g=0.473 a=0.527	G A	G A	G A
	A35	SNP	1.98198	1 / g	g=0.505	A	G A	G
	A53	SNP	1.12245	1 / 2g	g=0.445	G A	G	G A
	cumulative LR		17.426e6					

APPENDIX G: DNA.VIEW calculation of 16 STR markers (Powerplex® System) for tested samples..

Case	1	Scenario	1				
	M	Mother	1000-00186				
	C	Child	1000-00187				
	F	Tested Man	1000-00188				
C: M + F							
/C : M + ?							
	______						
	Malay	cumulative LR	1.09112e6	Posterior probability=	99.99991%	assuming prior=	50%
	______						
►	Malay				M	C	F
	D3S1358	2.01004	1 / 2q	q=0.249	16 17	17	16 17
	VWA	2.33333	1 / (p+q)	p=0.149 q=0.28	16 17	16 17	16 17
	D16S539	2.06818	1 / 2q	q=0.242	11 12	12	11 12
	D8S1179	2.96154	1 / 2t	t=0.169	11 15	15	12 15
	D21S11	2.19518	1 / 2q	q=0.228	29 31.2	29 30	30 34
	D18S51	1.91031	1 / 2r	r=0.262	14 16	14 15	13 15
	TH01	1.10486	1 / (2p+2s)	p=0.104 s=0.349	6 9	6 9	7 9
	FGA	4.20588	1 / 2c	c=0.119	21.2 22	22 24	23 24
	D7S820	3.06116	1 / s	s=0.327	8 11	11	11
	CSF1PO	1.52591	1 / 2q	q=0.328	11	11	10 11
	D13S317	1.99402	1 / 2s	s=0.251	8 11	11	8 11
	TPOX	1.49626	1 / (p+q)	p=0.536 q=0.132	8 9	8 9	8
	D5S818	4.13636	1 / q	q=0.242	10 11	11	11
	Penta E	16.2692	1 / s	s=0.0615	13 22	16 22	16
	Penta D	2.07831	1 / 2s	s=0.241	9 14	9 12	10 12
	cumulative LR	1.09112e6					

Case	2	Scenario	1				
	M	Mother	1000-00189				
	C	Child	1000-00190				
C: M + ?							
/C : ? + ?							
	Malay	cumulative LR	179401	Posterior probability=	99.9994%	assuming prior=	50%

►	Malay				M	C
	D3S1358	6.42137	(p+r) / 4pr	p=0.044 r=0.341	14 16	14 16
	VWA	1.15323	1 / 4s	s=0.217	17 18	15 18
	D16S539	1.84007	1 / 2p	p=0.272	11	11 12
	D8S1179	5.92308	1 / p	p=0.169	15	15
	D21S11	9.26852	1 / p	p=0.108	31	31
	D18S51	1.2703	1 / 4p	p=0.197	14 23	14 24
	TH01	2.70316	(p+q) / 4pq	p=0.126 q=0.349	8 9	8 9
	FGA	3.25	1 / 4p	p=0.0769	19 22	19 23
	D7S820	1.80036	1 / 4t	t=0.139	11 12	8 12
	CSF1PO	0.697075	1 / 4s	s=0.359	10 12	9 12
	D13S317	1.59395	1 / 4t	t=0.157	8 12	10 12
	TPOX	1.86406	1 / p	p=0.536	8	8
	D5S818	1.08804	1 / 4r	r=0.23	11 12	10 12
	Penta E	2.51786	1 / 4q	q=0.0993	10 11	11 15
	Penta D	2.10366	1 / 4q	q=0.119	9 10	10 12
	cumulative LR	179401				

Case	3	Scenario	1
	M	Mother	1000-00223
	C	Child	1000-00224
	F	Tested Man	1000-00225

C: M + F

/C: M + ?

Malay	cumulative LR	86.551e6	Posterior probability=	99.999999%	assuming prior=	50%

	Malay				M	C	F
►	Malay				M	C	F
	D3S1358	8.62931	1 / 2s	s=0.0579	15 16	15 18	17 18
	VWA	3.575	1 / r	r=0.28	15 18	15 17	17
	D16S539	2.52778	1 / (p+r)	p=0.272 r=0.124	11 13	11 13	11
	D8S1179	10.1111	1 / p	p=0.0989	12 15	10 15	10
	D21S11	3.76316	1 / 2a	a=0.133	29 30	30 32.2	32.2 34.2
	D18S51	2.89306	1 / 2s	s=0.173	13 15	13 16	15 16
	TH01	1.72586	1 / 2p	p=0.29	8 9	7 9	7 9
	FGA	2.31713	1 / 2p	p=0.216	25 26	22 25	22 23
	D7S820	1.53058	1 / 2r	r=0.327	9 13	11 13	10 11
	CSF1PO	1.39415	1 / 2r	r=0.359	12	12	10 12
	D13S317	1.91396	1 / (p+s)	p=0.272 s=0.251	8 11	8 11	8
	TPOX	20.02	1 / 2r	r=0.025	8 11	10 11	10 11
	D5S818	1.60417	1 / 2p	p=0.312	11	10 11	10 12
	Penta E	2.89726	1 / 2p	p=0.173	16 18	7 16	7 13
	Penta D	6.63462	1 / r	r=0.151	9	9 11	11
	cumulative LR	86.551e6					

Case	4	Scenario	1				
	M	Mother	1000-00194				
	C	Child	1000-00195				
	F	Tested Man	1000-00196				
C: M + F							
/C: M + ?							
	Malay	cumulative LR	43.4064e6	Posterior probability=	99.999998%	assuming prior=	50%
►	Malay				M	C	F
	D3S1358	1.46774	1 / 2p	p=0.341	17	16 17	16 17
	VWA	1.7875	1 / 2p	p=0.28	17 20	17	17 21
	D16S539	2.28539	1 / (p+q)	p=0.166 q=0.272	10 11	10 11	10
	D8S1179	10.1111	1 / p	p=0.0989	10 14	10	10
	D21S11	1.97047	1 / 2p	p=0.254	31 34.2	29 34.2	29 31
	D18S51	4.67757	1 / (2p+2x)	p=0.0939 x=0.013	13 21	13 21	13 16
	TH01	1.72586	1 / 2q	q=0.29	6 10	6 7	7 10
	FGA	2.90988	1 / 2p	p=0.172	24 25.2	21 24	21 22
	D7S820	1.76232	1 / (p+s)	p=0.241 s=0.327	8 11	8 11	11
	CSF1PO	3.4878	1 / (p+s)	p=0.219 s=0.0679	10 13	10 13	10
	D13S317	1.99402	1 / 2s	s=0.251	9 12	11 12	8 11
	TPOX	1.75614	1 / 2s	s=0.285	8	8 11	8 11
	D5S818	2.06818	1 / 2r	r=0.242	12	11 12	9 11
	Penta E	60.4286	1 / u	u=0.0165	14 16	16 19	19
	Penta D	5.75	1 / p	p=0.174	10 11	9 11	9
	cumulative LR	43.4064e6					

Case	5	Scenario	1
	M	Mother	1000-00197
	C	Child	1000-00198
	F	Tested Man	1000-00199

C: M + F

/C: M + ?

Malay	cumulative LR	87851	Posterior probability=	99.999%	assuming prior=	50%

► Malay				M	C	F
D3S1358	11.375	1 / 2p	p=0.044	15 18	14 15	14 15
VWA	3.35906	1 / 2q	q=0.149	17 18	16 18	15 16
D16S539	2.52778	1 / (p+r)	p=0.272 r=0.124	11 13	11 13	11 13
D8S1179	4.46875	1 / 2p	p=0.112	15 16	11 15	11 15
D21S11	11.6395	1 / 2a	a=0.043	29	29 33.2	31 33.2
D18S51	5.32447	1 / 2p	p=0.0939	15 21	13 21	13 14
TH01	1.4341	1 / 2s	s=0.349	7 9	9	6 9
FGA	1.70819	1 / (2p+2s)	p=0.0769 s=0.216	19 22	19 22	20 22
D7S820	1.76232	1 / (p+s)	p=0.241 s=0.327	8 11	8 11	11
CSF1PO	0.0022	μ	μ=0.0022	11	11	10 12
D13S317	11.1222	1 / 2r	r=0.045	12	12 13	11 13
TPOX	1.86406	1 / p	p=0.536	9	8 9	8
D5S818	2.06818	1 / 2r	r=0.242	9 10	9 11	11 12
Penta E	2.43103	1 / 2p	p=0.206	15 16	12 15	12 13
Penta D	3.31731	1 / 2q	q=0.151	10 12	10 11	11 12
cumulative LR	87851					

Case	6	Scenario	1				
	M	Mother	1000-00200				
	C	Child	1000-00201				
	F	Tested Man	1000-00202				
C: M + F							
/C: M + ?							
	Malay	cumulative LR	11.407e9	Posterior probability=	99.99999999%	assuming prior=	50%
►	Malay				M	C	F
	D3S1358	2.01004	1 / 2r	r=0.249	15 18	17 18	16 17
	VWA	3.35906	1 / 2r	r=0.149	14 17	16 17	16 17
	D16S539	4.03629	1 / 2r	r=0.124	11 13	13	12 13
	D8S1179	2.42961	1 / 2r	r=0.206	11 15	13 15	13 15
	D21S11	4.63426	1 / 2q	q=0.108	30 33.2	31 33.2	31 32
	D18S51	7.15	1 / 2t	t=0.0699	13 18	17 18	15 17
	TH01	3.97222	1 / 2q	q=0.126	7 8	8	7 8
	FGA	100.1	1 / 2p	p=0.005	22 23	18 23	18 24
	D7S820	4.15353	1 / p	p=0.241	8 12	8	8
	CSF1PO	4.57078	1 / p	p=0.219	12	10 12	10
	D13S317	3.98805	1 / s	s=0.251	8 13	8 11	11
	TPOX	3.79167	1 / 2q	q=0.132	9	9	8 9
	D5S818	2.17609	1 / 2q	q=0.23	11 12	12	11 12
	Penta E	10.0714	1 / p	p=0.0993	11 16	11	11
	Penta D	2.07831	1 / 2s	s=0.241	9	9 12	9 12
	cumulative LR	11.407e9					

Case	7	Scenario	1				
	M	Mother	1000-00203				
	C	Child	1000-00204				
C: M + ?							
/C: ? + ?							
	Malay	cumulative LR	108.978	Posterior probability=	99.09%	assuming prior=	50%

►	Malay				M	C
	D3S1358	1.68519	1 / 2p	p=0.297	15 16	15
	VWA	1.67953	1 / 4p	p=0.149	16 18	16 19
	D16S539	1.43822	1 / 4p	p=0.174	9 10	9 13
	D8S1179	1.31021	1 / 4t	t=0.191	14 16	10 14
	D21S11	0.985236	1 / 4q	q=0.254	28 29	29 32.2
	D18S51	1.44653	1 / 4r	r=0.173	14 16	16 19
	TH01	0.717049	1 / 4s	s=0.349	6 9	7 9
	FGA	1.15856	1 / 4p	p=0.216	22 23	22 24
	D7S820	1.80036	1 / 4t	t=0.139	8 12	10 12
	CSF1PO	0.762957	1 / 4r	r=0.328	9 11	11 12
	D13S317	0.997012	1 / 4r	r=0.251	9 11	10 11
	TPOX	1.75614	1 / 2s	s=0.285	8 11	11
	D5S818	1.83617	(p+q) / 4pq	p=0.312 q=0.242	10 11	10 11
	Penta E	1.44863	1 / 4p	p=0.173	7 16	7 19
	Penta D	2.69781	(p+q) / 4pq	p=0.151 q=0.241	11 12	11 12
	cumulative LR	108.978				

Case	8C	Scenario	1
	M	Mother	1000-00205
	C	Child #1	1000-00206
	D	Child #2	1000-00207
	E	Child #3	1000-00208
	F	Tested Man	1000-00209

C: M + F

/C: M + ?

Malay	cumulative LR	516.24e6	Posterior probability=	99.9999998%	assuming prior=	50%		
Malay				M	C	D	E	F
D3S1358	2.01004	1 / 2s	s=0.249	14 16	16 17	16	16	16 17
VWA	10.01	1 / 2q	q=0.05	15 17	15	14 17	14 15	14 15
D16S539	1.84007	1 / 2r	r=0.272	9 10	10 11	9 11	9 11	9 11
D8S1179	2.42961	1 / 2p	p=0.206	13	13	13	13 16	13 15
D21S11	11.6395	1 / 2a	a=0.043	30 31	31 33.2	30 33.2	28 31	28 33.2
D18S51	2.89306	1 / 2t	t=0.173	15 19	16 19	12 15	16 19	12 16
TH01	8.20492	1 / 2a	a=0.0609	9	9 9.3	9	9 9.3	9 9.3
FGA	2.90988	1 / 2p	p=0.172	23 27	21 23	23 24	25 27	21 24
D7S820	2.54061	1 / 2p	p=0.197	11 12	10 12	10 11	10 12	10 12
CSF1PO	3.05183	1 / q	q=0.328	10 11	11	10 11	11	11
D13S317	11.1222	1 / 2u	u=0.045	8 11	8 13	11 13	8 13	8 13
TPOX	1.21776	1 / (p+s)	p=0.536 s=0.285	8 11	8 11	8 11	8 11	8 11
D5S818	1.60417	1 / 2p	p=0.312	10 13	10	10	10	10 13
Penta E	5.03571	1 / 2p	p=0.0993	14 19	11 19	11 19	12 19	11 12
Penta D	8.41463	1 / q	q=0.119	9 11	9 10	9 10	10 11	10
cumulative LR	516.24e6							

Case	8D	Scenario	1
	M	Mother	1000-00205
	C	Child #1	1000-00206
	D	Child #2	1000-00207
	E	Child #3	1000-00208
	F	Tested Man	1000-00209

D: M + F

/D: M + ?

			Posterior		assuming	
Malay	cumulative LR	27.719e6	probability=	99.999996%	prior=	50%

► Malay				M	C	D	E	F
D3S1358	1.46774	1 / 2r	r=0.341	14 16	16 17	16	16	16 17
VWA	2.52778	1 / 2p	p=0.198	15 17	15	14 17	14 15	14 15
D16S539	1.84007	1 / 2r	r=0.272	9 10	10 11	9 11	9 11	9 11
D8S1179	2.42961	1 / 2p	p=0.206	13	13	13	13 16	13 15
D21S11	11.6395	1 / 2a	a=0.043	30 31	31 33.2	30 33.2	28 31	28 33.2
D18S51	5.56111	1 / 2p	p=0.0899	15 19	16 19	12 15	16 19	12 16
TH01	1.4341	1 / 2p	p=0.349	9	9 9.3	9	9 9.3	9 9.3
FGA	4.20588	1 / 2s	s=0.119	23 27	21 23	23 24	25 27	21 24
D7S820	2.54061	1 / 2p	p=0.197	11 12	10 12	10 11	10 12	10 12
CSF1PO	1.82998	1 / (p+q)	p=0.219 q=0.328	10 11	11	10 11	11	11
D13S317	11.1222	1 / 2u	u=0.045	8 11	8 13	11 13	8 13	8 13
TPOX	1.21776	1 / (p+s)	p=0.536 s=0.285	8 11	8 11	8 11	8 11	8 11
D5S818	1.60417	1 / 2p	p=0.312	10 13	10	10	10	10 13
Penta E	5.03571	1 / 2p	p=0.0993	14 19	11 19	11 19	12 19	11 12
Penta D	8.41463	1 / q	q=0.119	9 11	9 10	9 10	10 11	10
cumulative LR	27.719e6							

Case	8E	Scenario	1
	M	Mother	1000-00205
	C	Child #1	1000-00206
	D	Child #2	1000-00207
	E	Child #3	1000-00208
		Tested	
	F	Man	1000-00209

E: M + F

/E: M + ?

	cumulative		Posterior		assuming	
Malay	LR	54.4425	probability=	98.2%	prior=	50%

► Malay				M	C	D	E	F
D3S1358	1.46774	1 / 2r	r=0.341	14 16	16 17	16	16	16 17
VWA	2.52778	1 / 2p	p=0.198	15 17	15	14 17	14 15	14 15
D16S539	1.84007	1 / 2r	r=0.272	9 10	10 11	9 11	9 11	9 11
D8S1179	0.0028	µ	µ=0.0028	13	13	13	13 16	13 15
D21S11	7.25362	1 / 2p	p=0.0689	30 31	31 33.2	30 33.2	28 31	28 33.2
D18S51	2.89306	1 / 2t	t=0.173	15 19	16 19	12 15	16 19	12 16
TH01	8.20492	1 / 2a	a=0.0609	9	9 9.3	9	9 9.3	9 9.3
FGA	0.0048	µ	µ=0.0048	23 27	21 23	23 24	25 27	21 24
D7S820	2.54061	1 / 2p	p=0.197	11 12	10 12	10 11	10 12	10 12
CSF1PO	3.05183	1 / q	q=0.328	10 11	11	10 11	11	11
D13S317	11.1222	1 / 2u	u=0.045	8 11	8 13	11 13	8 13	8 13
TPOX	1.21776	1 / (p+s)	p=0.536 s=0.285	8 11	8 11	8 11	8 11	8 11
D5S818	1.60417	1 / 2p	p=0.312	10 13	10	10	10	10 13
Penta E	2.43103	1 / 2q	q=0.206	14 19	11 19	11 19	12 19	11 12
Penta D	8.41463	1 / q	q=0.119	9 11	9 10	9 10	10 11	10
cumulative LR	54.4425							

Case	9G	Scenario	1
	M	Mother	1000-00210
	C	Child	1000-00211
	G	Man #2	1000-00212

C: G + ?

/C: ? + ?

	Malay	cumulative LR	30.0776	Posterior probability=	96.8%	assuming prior=	50%

►	Malay				M	C	G
	D3S1358	1.46774	1 / 2p	p=0.341	16	16 17	16
	VWA	2.83276	(p+r) / 4pr	p=0.149 r=0.217	18	16 18	16 18
	D16S539	1.84007	1 / 2p	p=0.272	11 12	11 13	11
	D8S1179	1.48077	1 / 4t	t=0.169	11 14	11 15	14 15
	D21S11	0.0022	µ	µ=0.0022	28 31	28 31	29 32
	D18S51	7.37182	(p+s) / 4ps	p=0.262 s=0.039	15 18	15 18	15 18
	TH01	1.98611	1 / 4p	p=0.126	8 9	8 9	8 10
	FGA	2.61351	(p+q) / 4pq	p=0.172 q=0.216	21 22	21 22	21 22
	D7S820	1.2703	1 / 4r	r=0.197	10 12	8 10	10 12
	CSF1PO	1.39415	1 / 2q	q=0.359	12	12	11 12
	D13S317	3.82061	1 / 2p	p=0.131	9	9 11	9
	TPOX	0.466015	1 / 4p	p=0.536	9 10	8 9	8 11
	D5S818	2.06818	1 / 2p	p=0.242	11	11 12	11
	Penta E	2.29891	1 / 4r	r=0.109	11 13	11 13	13 16
	Penta D	2.10366	1 / 4p	p=0.119	10 11	10 12	10 11
	cumulative LR	30.0776					

Case	9M	Scenario	1				
	M	Mother	1000-00210				
	C	Child	1000-00211				
	G	Man #2	1000-00212				
C: M + ?							
/C: ? + ?							
	Malay	cumulative LR	1.15824e6	Posterior probability=	99.99991%	assuming prior=	50%
►	Malay				M	C	G
	D3S1358	1.46774	1 / 2p	p=0.341	16	16 17	16
	VWA	2.30645	1 / 2r	r=0.217	18	16 18	16 18
	D16S539	0.920037	1 / 4p	p=0.272	11 12	11 13	11
	D8S1179	2.23438	1 / 4p	p=0.112	11 14	11 15	14 15
	D21S11	5.94394	(p+s) / 4ps	p=0.0689 s=0.108	28 31	28 31	29 32
	D18S51	7.37182	(p+s) / 4ps	p=0.262 s=0.039	15 18	15 18	15 18
	TH01	2.70316	(p+q) / 4pq	p=0.126 q=0.349	8 9	8 9	8 10
	FGA	2.61351	(p+q) / 4pq	p=0.172 q=0.216	21 22	21 22	21 22
	D7S820	1.2703	1 / 4r	r=0.197	10 12	8 10	10 12
	CSF1PO	2.7883	1 / q	q=0.359	12	12	11 12
	D13S317	3.82061	1 / 2p	p=0.131	9	9 11	9
	TPOX	1.89583	1 / 4q	q=0.132	9 10	8 9	8 11
	D5S818	2.06818	1 / 2p	p=0.242	11	11 12	11
	Penta E	4.81677	(p+r) / 4pr	p=0.0993 r=0.109	11 13	11 13	13 16
	Penta D	2.10366	1 / 4p	p=0.119	10 11	10 12	10 11
	cumulative LR	1.15824e6					

Case	10	Scenario	1				
	C	Child	1000-00231				
	F	Tested Man	1000-00232				
C: ? + F							
/C: ? + ?							
	Malay	cumulative LR	275561	Posterior probability=	99.9996%	assuming prior=	50%
►	Malay				C	F	
	D3S1358	2.01004	1 / 2r	r=0.249	17	15 17	
	VWA	2.63421	1 / 4s	s=0.0949	18 19	16 19	
	D16S539	1.50753	1 / 4p	p=0.166	10 11	10 13	
	D8S1179	4.76667	1 / 2u	u=0.105	10 16	16	
	D21S11	6.76351	1 / 4a	a=0.037	30 30.2	30.2 31	
	D18S51	3.82061	1 / p	p=0.262	15	15	
	TH01	0.862931	1 / 4p	p=0.29	7 9	7 10	
	FGA	3.20833	1 / 2p	p=0.156	23	23 24	
	D7S820	3.06116	1 / p	p=0.327	11	11	
	CSF1PO	0.697075	1 / 4r	r=0.359	11 12	10 12	
	D13S317	0.997012	1 / 4s	s=0.251	8 11	10 11	
	TPOX	1.75614	1 / 2r	r=0.285	11	9 11	
	D5S818	1.60417	1 / 2p	p=0.312	10 11	10	
	Penta E	5.875	1 / 4t	t=0.0426	11 15	15 20	
	Penta D	2.875	1 / 2p	p=0.174	9 10	9	
	cumulative LR	275561					

Case	11	Scenario	1				
	M	Mother	1000-00215				
	C	Child	1000-00216				

C: M + ?

/C: ? + ?

Malay	cumulative LR	5.2374e6	Posterior probability=	99.99998%	assuming prior=	50%

► Malay				M	C
D3S1358	5.31968	(p+q) / 4pq	p=0.249 q=0.0579	17 18	17 18
VWA	5.005	1 / 4q	q=0.05	14 15	15 19
D16S539	1.50753	1 / 4q	q=0.166	10 12	9 10
D8S1179	2.52778	1 / 4p	p=0.0989	10 16	10 15
D21S11	7.15	1 / 4u	u=0.035	31 32	27 32
D18S51	5.08122	1 / p	p=0.197	14	14
TH01	3.45172	1 / p	p=0.29	7	7
FGA	1.60417	1 / 4b	b=0.156	22 23	21.2 23
D7S820	3.47569	1 / 4q	q=0.0719	8 9	9 10
CSF1PO	0.762957	1 / 4r	r=0.328	9 11	11 14
D13S317	0.920037	1 / 4p	p=0.272	8 10	8 11
TPOX	0.93203	1 / 2q	q=0.536	8	7 8
D5S818	2.59815	(p+r) / 4pr	p=0.242 r=0.16	11 13	11 13
Penta E	7.55357	1 / 2p	p=0.0662	5	5 12
Penta D	5.75	1 / p	p=0.174	9	9
cumulative LR	5.2374e6				

Case	12	Scenario	1				
	M	Mother	1000-00233				
	C	Child	1000-00234				
	F	Tested Man	1000-00235				
C: M + F							
/C: M + ?							
	Malay	cumulative LR	193.895e6	Posterior probability=	99.9999995%	assuming prior=	50%
►	Malay				M	C	F
	D3S1358	4.02008	1 / q	q=0.249	16 17	17	17
	VWA	1.7875	1 / 2s	s=0.28	14 17	17	16 17
	D16S539	1.84007	1 / 2q	q=0.272	10 11	11	11 13
	D8S1179	2.42961	1 / 2s	s=0.206	10 14	13 14	13 16
	D21S11	14.3	1 / 2s	s=0.035	29	29 32	32 32.2
	D18S51	5.32447	1 / 2p	p=0.0939	14 19	13 19	13 14
	TH01	3.45172	1 / p	p=0.29	9	7 9	7
	FGA	4.20588	1 / 2q	q=0.119	23 26	24 26	23 24
	D7S820	3.60072	1 / 2t	t=0.139	11	11 12	8 12
	CSF1PO	1.39415	1 / 2r	r=0.359	11	11 12	10 12
	D13S317	3.98805	1 / q	q=0.251	10 11	11	11
	TPOX	1.86406	1 / p	p=0.536	10 11	8 11	8
	D5S818	6.25625	1 / s	s=0.16	10 13	13	13
	Penta E	8.13462	1 / 2u	u=0.0615	11 16	16	14 16
	Penta D	2.875	1 / 2p	p=0.174	10 13	9 10	9 10
	cumulative LR	193.895e6					

CONFERENCES/PROCEEDINGS

Genotyping of 52 Autosomal SNPs (SNPforID) in Malaysian Population. 19th World Meeting of The International Association of Forensic Sciences (IAFS), 12th to 17th September 2011 at Funchal, Madeira, Portugal.

Application of Single Nucleotide Polymorphisms (SNPs) to Forensic Casework in Malaysia. American Academy of Forensic Sciences, 21st to 25th February, 2012 at Atlanta, Georgia, USA.

Application of Single Nucleotide Polymorphisms (SNPs) To Forensic Casework in Malaysia. Postgraduate Research Conference, University of Central Lancashire, 2nd to 5th July, 2012.

SNPs Genotyping of Forensic Casework Samples using 52 SNPforID markers. ISFG 2013, Melbourne Convention Centre, Australia, 2nd to 7th September, 2013.

Application of 52 Autosomal SNPs (SNPforID) to Forensic Casework in Malaysia. INPALMS 2013, Kuala Lumpur, Malaysia, 5th to 10th October, 2013.

PUBLICATIONS

S. Alimat, S. Hadi, W. Goodwin (2013). *SNPs Genotyping of Forensic Casework Samples using the 52 SNPforID markers*. Forensic Science International:Genetics Supplement Series, Vol 4, Issue 1, 2013, pg 178-179.

M.S.Nazir, S.Iyavoo, S.Alimat, N.Zahra, S.H.Sanqoor, J.A. Smith, C.Moffatt, W.Goodwin (2013). *Development of a multiplex system to assess DNA persistence in taphonomic studies*. Electrophoresis (December 2013), Volume 34, Issue 24, pg 3352-3360.

EVIDENCE PRESENTATIONS AT THE MALAYSIAN COURTS (HIGH COURT AND SESSIONS) AS AN EXPERT WITNESS – 7 CASES (within the following dates):

1. 01/12/2009 to 06/12/2009
2. 26/02/2010 to 07/03/2010
3. 08/06/2010 to 18/06/2010
4. 28/04/2011 to 09/05/2011

REFERENCES

ABDULLA, M.A., AHMED, I., ASSAWAMAKIN, A., BHAK, J., BRAHMACHARI, S.K., CALACAL, G.C., CHAURASIA, A., CHEN, C., CHEN, J., CHEN, Y., CHU, J., Cutiongco-de la Paz,Eva Maria C., DE UNGRIA, M.C.A., DELFIN, F.C., EDO, J., FUCHAREON, S., GHANG, H., GOJOBORI, T., HAN, J., HO, S., HOH, B.P., HUANG, W., INOKO, H., JHA, P., JINAM, T.A., JIN, L., JUNG, J., KANGWANPONG, D., KAMPUANSAI, J., KENNEDY, G.C., KHURANA, P., KIM, H., KIM, K., KIM, S., KIM, W., KIMM, K., KIMURA, R., KOIKE, T., KULAWONGANUNCHAI, S., KUMAR, V., LAI, P.S., LEE, J., LEE, S., LIU, E.T., MAJUMDER, P.P., MANDAPATI, K.K., MARZUKI, S., MITCHELL, W., MUKERJI, M., NARITOMI, K., NGAMPHIW, C., NIIKAWA, N., NISHIDA, N., OH, B., OH, S., OHASHI, J., OKA, A., ONG, R., PADILLA, C.D., PALITTAPONGARNPIM, P., PERDIGON, H.B., PHIPPS, M.E., PNG, E., SAKAKI, Y., SALVADOR, J.M., SANDRALING, Y., SCARIA, V., SEIELSTAD, M., SIDEK, M.R., SINHA, A., SRIKUMMOOL, M., SUDOYO, H., SUGANO, S., SURYADI, H., SUZUKI, Y., TABBADA, K.A., TAN, A., TOKUNAGA, K., TONGSIMA, S., VILLAMOR, L.P., WANG, E., WANG, Y., WANG, H., WU, J., XIAO, H., XU, S., YANG, J.O., SHUGART, Y.Y., YOO, H., YUAN, W., ZHAO, G., ZILFALIL, B.A., HUGO Pan Asian SNP Consortium and Indian Genome Variation Consortium, 2009. Mapping Human Genetic Diversity in Asia. *Science*, DEC 11 2009, vol. 326, no. 5959, pp. 1541-1545 ISSN 0036-8075. DOI 10.1126/science.1177074.

ALAEDDINI, R., WALSH, S.J. and ABBAS, A., 2010. Forensic implications of genetic analyses from degraded DNA-A review. *Forensic Science International-Genetics*, APR 2010, vol. 4, no. 3, pp. 148-157 ISSN 1872-4973. DOI 10.1016/j.fsigen.2009.09.007.

ALBARGHOUTHI, M.N., BUCHHOLZ, B.A., HUIBERTS, P.J., STEIN, T.M. and BARRON, A.E., 2002. Poly-N-hydroxyethylacrylamide (polyDuramide?): A novel, hydrophilic, self-coating polymer matrix for DNA sequencing by capillary electrophoresis. *Electrophoresis*, vol. 23, no. 10, pp. 1429-1440 ISSN 1522-2683. DOI 10.1002/1522-2683(200205)23:10<1429::AID-ELPS1429>3.0.CO;2-A.

ALTSHULER, D.M., DURBIN, R.M., ABECASIS, G.R., BENTLEY, D.R., CHAKRAVARTI, A., CLARK, A.G., DONNELLY, P., EICHLER, E.E., FLICEK, P., GABRIEL, S.B., GIBBS, R.A., GREEN, E.D., HURLES, M.E., KNOPPERS, B.M., KORBEL, J.O., LANDER, E.S., LEE, C., LEHRACH, H., MARDIS, E.R., MARTH, G.T., MCVEAN, G.A., NICKERSON, D.A., SCHMIDT, J.P., SHERRY, S.T., WANG, J., WILSON, R.K., GIBBS, R.A., DINH, H., KOVAR, C., LEE, S., LEWIS, L., MUZNY, D., REID, J., WANG, M., WANG, J., FANG, X., GUO, X., JIAN, M., JIANG, H., JIN, X., LI, G., LI, J., LI, Y., LI, Z., LIU, X., LU, Y., MA, X., SU, Z., TAI, S., TANG, M., WANG, B., WANG, G., WU, H., WU, R., YIN, Y., ZHANG, W., ZHAO, J., ZHAO, M., ZHENG, X., ZHOU, Y., LANDER, E.S., ALTSHULER, D.M., GABRIEL, S.B., GUPTA, N., FLICEK, P., CLARKE, L., LEINONEN, R., SMITH, R.E., ZHENG-BRADLEY, X., BENTLEY, D.R., GROCOCK, R., HUMPHRAY, S., JAMES, T., KINGSBURY, Z., LEHRACH, H., SUDBRAK, R., ALBRECHT, M.W., AMSTISLAVSKIY, V.S., BORODINA, T.A., LIENHARD, M., MERTES, F., SULTAN, M., TIMMERMANN, B., YASPO, M., SHERRY, S.T., MCVEAN, G.A., MARDIS, E.R., WILSON, R.K., FULTON, L., FULTON, R., WEINSTOCK, G.M., DURBIN, R.M., BALASUBRAMANIAM, S., BURTON, J., DANECEK, P., KEANE, T.M., KOLB-KOKOCINSKI, A., MCCARTHY, S., STALKER, J., QUAIL, M., SCHMIDT, J.P., DAVIES, C.J., GOLLUB, J., WEBSTER, T., WONG, B., ZHAN, Y., AUTON, A., GIBBS, R.A., YU, F., BAINBRIDGE, M., CHALLIS, D., EVANI, U.S., LU, J., MUZNY, D., NAGASWAMY, U., REID, J., SABO, A., WANG, Y., YU, J., WANG, J., COIN, L.J.M., FANG, L., GUO, X., JIN, X., LI, G., LI, Q., LI, Y., LI, Z., LIN, H., LIU, B., LUO, R., QIN, N., SHAO, H., WANG, B., XIE, Y., YE, C., YU, C., ZHANG, F., ZHENG,

H., ZHU, H., MARTH, G.T., GARRISON, E.P., KURAL, D., LEE, W., LEONG, W.F., WARD, A.N., WU, J., ZHANG, M., LEE, C., GRIFFIN, L., HSIEH, C., MILLS, R.E., SHI, X., VON GROTTHUSS, M., ZHANG, C., DALY, M.J., DEPRISTO, M.A., ALTSHULER, D.M., BANKS, E., BHATIA, G., CARNEIRO, M.O., DEL ANGEL, G., GABRIEL, S.B., GENOVESE, G., GUPTA, N., HANDSAKER, R.E., HARTL, C., LANDER, E.S., MCCARROLL, S.A., NEMESH, J.C., POPLIN, R.E., SCHAFFNER, S.F., SHAKIR, K., YOON, S.C., LIHM, J., MAKAROV, V., JIN, H., KIM, W., KIM, K.C., KORBEL, J.O., RAUSCH, T., FLICEK, P., BEAL, K., CLARKE, L., CUNNINGHAM, F., HERRERO, J., MCLAREN, W.M., RITCHIE, G.R.S., SMITH, R.E., ZHENG-BRADLEY, X., CLARK, A.G., GOTTIPATI, S., KEINAN, A., RODRIGUEZ-FLORES, J.L., SABETI, P.C., GROSSMAN, S.R., TABRIZI, S., TARIYAL, R., COOPER, D.N., BALL, E.V., STENSON, P.D., BENTLEY, D.R., BARNES, B., BAUER, M., CHEETHAM, R.K., COX, T., EBERLE, M., HUMPHRAY, S., KAHN, S., MURRAY, L., PEDEN, J., SHAW, R., YE, K., BATZER, M.A., KONKEL, M.K., WALKER, J.A., MACARTHUR, D.G., LEK, M., SUDBRAK, R., AMSTISLAVSKIY, V.S., HERWIG, R., SHRIVER, M.D., BUSTAMANTE, C.D., BYRNES, J.K., DE LA VEGA, F.M., GRAVEL, S., KENNY, E.E., KIDD, J.M., LACROUTE, P., MAPLES, B.K., MORENO-ESTRADA, A., ZAKHARIA, F., HALPERIN, E., BARAN, Y., CRAIG, D.W., CHRISTOFORIDES, A., HOMER, N., IZATT, T., KURDOGLU, A.A., SINARI, S.A., SQUIRE, K., SHERRY, S.T., XIAO, C., SEBAT, J., BAFNA, V., YE, K., BURCHARD, E.G., HERNANDEZ, R.D., GIGNOUX, C.R., HAUSSLER, D., KATZMAN, S.J., KENT, W.J., HOWIE, B., RUIZ-LINARES, A., DERMITZAKIS, E.T., LAPPALAINEN, T., DEVINE, S.E., LIU, X., MAROO, A., TALLON, L.J., ROSENFELD, J.A., MICHELSON, L.P., ABECASIS, G.R., KANG, H.M., ANDERSON, P., ANGIUS, A., BIGHAM, A., BLACKWELL, T., BUSONERO, F., CUCCA, F., FUCHSBERGER, C., JONES, C., JUN, G., LI, Y., LYONS, R., MASCHIO, A., PORCU, E., REINIER, F., SANNA, S., SCHLESSINGER, D., SIDORE, C., TAN, A., TROST, M.K., AWADALLA, P., HODGKINSON, A., LUNTER, G., MCVEAN, G.A., MARCHINI, J.L., MYERS, S., CHURCHHOUSE, C., DELANEAU, O., GUPTA-HINCH, A., IQBAL, Z., MATHIESON, I., RIMMER, A., XIFARA, D.K., OLEKSYK, T.K., FU, Y., LIU, X., XIONG, M., JORDE, L., WITHERSPOON, D., XING, J., EICHLER, E.E., BROWNING, B.L., ALKAN, C., HAJIRASOULIHA, I., HORMOZDIARI, F., KO, A., SUDMANT, P.H., MARDIS, E.R., CHEN, K., CHINWALLA, A., DING, L., DOOLING, D., KOBOLDT, D.C., MCLELLAN, M.D., WALLIS, J.W., WENDL, M.C., ZHANG, Q., DURBIN, R.M., HURLES, M.E., TYLER-SMITH, C., ALBERS, C.A., AYUB, Q., BALASUBRAMANIAM, S., CHEN, Y., COFFEY, A.J., COLONNA, V., DANECEK, P., HUANG, N., JOSTINS, L., KEANE, T.M., LI, H., MCCARTHY, S., SCALLY, A., STALKER, J., WALTER, K., XUE, Y., ZHANG, Y., GERSTEIN, M.B., ABYZOV, A., BALASUBRAMANIAN, S., CHEN, J., CLARKE, D., FU, Y., HABEGGER, L., HARMANCI, A.O., JIN, M., KHURANA, E., MU, X.J., SISU, C., LI, Y., LUO, R., ZHU, H., LEE, C., GRIFFIN, L., HSIEH, C., MILLS, R.E., SHI, X., VON GROTTHUSS, M., ZHANG, C., MARTH, G.T., GARRISON, E.P., KURAL, D., LEE, W., WARD, A.N., WU, J., ZHANG, M., MCCARROLL, S.A., ALTSHULER, D.M., BANKS, E., DEL ANGEL, G., GENOVESE, G., HANDSAKER, R.E., HARTL, C., NEMESH, J.C., SHAKIR, K., YOON, S.C., LIHM, J., MAKAROV, V., DEGENHARDT, J., FLICEK, P., CLARKE, L., SMITH, R.E., ZHENG-BRADLEY, X., KORBEL, J.O., RAUSCH, T., STUETZ, A.M., BENTLEY, D.R., BARNES, B., CHEETHAM, R.K., EBERLE, M., HUMPHRAY, S., KAHN, S., MURRAY, L., SHAW, R., YE, K., BATZER, M.A., KONKEL, M.K., WALKER, J.A., LACROUTE, P., CRAIG, D.W., HOMER, N., CHURCH, D., XIAO, C., SEBAT, J., BAFNA, V., MICHAELSON, J.J., YE, K., DEVINE, S.E., LIU, X., MAROO, A., TALLON, L.J., LUNTER, G., MCVEAN, G.A., IQBAL, Z., WITHERSPOON, D., XING, J., EICHLER, E.E., ALKAN, C., HAJIRASOULIHA, I., HORMOZDIARI, F., KO, A., SUDMANT, P.H., CHEN, K., CHINWALLA, A., DING, L., MCLELLAN, M.D., WALLIS, J.W., HURLES, M.E., Ben Blackburne, LI, H., LINDSAY, S.J., NING, Z., SCALLY, A., WALTER, K., ZHANG, Y., GERSTEIN, M.B., ABYZOV, A., CHEN, J., CLARKE, D., KHURANA, E., MU, X.J., SISU, C., GIBBS, R.A., YU, F., BAINBRIDGE, M., CHALLIS, D., EVANI, U.S., KOVAR, C., LEWIS, L., LU, J., MUZNY, D., NAGASWAMY, U., REID, J., SABO, A., YU, J., GUO, X., LI, Y., WU, R., MARTH, G.T., GARRISON, E.P., LEONG, W.F., WARD, A.N., DEL ANGEL, G., DEPRISTO, M.A., GABRIEL, S.B., GUPTA, N., HARTL, C., POPLIN, R.E., CLARK, A.G., RODRIGUEZ-FLORES, J.L., FLICEK, P., CLARKE, L., SMITH, R.E., ZHENG-BRADLEY, X., MACARTHUR, D.G., BUSTAMANTE, C.D., GRAVEL, S., CRAIG, D.W., CHRISTOFORIDES, A., HOMER, N., IZATT, T., SHERRY, S.T., XIAO,

C., DERMITZAKIS, E.T., ABECASIS, G.R., KANG, H.M., MCVEAN, G.A., MARDIS, E.R., DOOLING, D., FULTON, L., FULTON, R., KOBOLDT, D.C., DURBIN, R.M., BALASUBRAMANIAM, S., KEANE, T.M., MCCARTHY, S., STALKER, J., GERSTEIN, M.B., BALASUBRAMANIAN, S., HABEGGER, L., GARRISON, E.P., GIBBS, R.A., BAINBRIDGE, M., MUZNY, D., YU, F., YU, J., DEL ANGEL, G., HANDSAKER, R.E., MAKAROV, V., RODRIGUEZ-FLORES, J.L., JIN, H., KIM, W., KIM, K.C., FLICEK, P., BEAL, K., CLARKE, L., CUNNINGHAM, F., HERRERO, J., MCLAREN, W.M., RITCHIE, G.R.S., ZHENG-BRADLEY, X., TABRIZI, S., MACARTHUR, D.G., LEK, M., BUSTAMANTE, C.D., DE LA VEGA, F.M., CRAIG, D.W., KURDOGLU, A.A., LAPPALAINEN, T., ROSENFELD, J.A., MICHELSON, L.P., AWADALLA, P., HODGKINSON, A., MCVEAN, G.A., CHEN, K., TYLER-SMITH, C., CHEN, Y., COLONNA, V., FRANKISH, A., HARROW, J., XUE, Y., GERSTEIN, M.B., ABYZOV, A., BALASUBRAMANIAN, S., CHEN, J., CLARKE, D., FU, Y., HARMANCI, A.O., JIN, M., KHURANA, E., MU, X.J., SISU, C., GIBBS, R.A., FOWLER, G., HALE, W., KALRA, D., KOVAR, C., MUZNY, D., REID, J., WANG, J., GUO, X., LI, G., LI, Y., ZHENG, X., ALTSHULER, D.M., FLICEK, P., CLARKE, L., BARKER, J., KELMAN, G., KULESHA, E., LEINONEN, R., MCLAREN, W.M., RADHAKRISHNAN, R., ROA, A., SMIRNOV, D., SMITH, R.E., STREETER, I., TONEVA, I., VAUGHAN, B., ZHENG-BRADLEY, X., BENTLEY, D.R., COX, T., HUMPHRAY, S., KAHN, S., SUDBRAK, R., ALBRECHT, M.W., LIENHARD, M., CRAIG, D.W., IZATT, T., KURDOGLU, A.A., SHERRY, S.T., ANANIEV, V., BELAIA, Z., BELOSLYUDTSEV, D., BOUK, N., CHEN, C., CHURCH, D., COHEN, R., COOK, C., GARNER, J., HEFFERON, T., KIMELMAN, M., LIU, C., LOPEZ, J., MERIC, P., O'SULLIVAN, C., OSTAPCHUK, Y., PHAN, L., PONOMAROV, S., SCHNEIDER, V., SHEKHTMAN, E., SIROTKIN, K., SLOTTA, D., XIAO, C., ZHANG, H., HAUSSLER, D., ABECASIS, G.R., MCVEAN, G.A., ALKAN, C., KO, A., DOOLING, D., DURBIN, R.M., BALASUBRAMANIAM, S., KEANE, T.M., MCCARTHY, S., STALKER, J., CHAKRAVARTI, A., KNOPPERS, B.M., ABECASIS, G.R., BARNES, K.C., BEISWANGER, C., BURCHARD, E.G., BUSTAMANTE, C.D., CAI, H., CAO, H., DURBIN, R.M., GHARANI, N., GIBBS, R.A., GIGNOUX, C.R., GRAVEL, S., HENN, B., JONES, D., JORDE, L., KAYE, J.S., KEINAN, A., KENT, A., KERASIDOU, A., LI, Y., MATHIAS, R., MCVEAN, G.A., MORENO-ESTRADA, A., OSSORIO, P.N., PARKER, M., REICH, D., ROTIMI, C.N., ROYAL, C.D., SANDOVAL, K., SU, Y., SUDBRAK, R., TIAN, Z., TIMMERMANN, B., TISHKOFF, S., TOJI, L.H., TYLER-SMITH, C., VIA, M., WANG, Y., YANG, H., YANG, L., ZHU, J., BODMER, W., BEDOYA, G., RUIZ-LINARES, A., MING, C.Z., YANG, G., YOU, C.J., PELTONEN, L., GARCIA-MONTERO, A., ORFAO, A., DUTIL, J., MARTINEZ-CRUZADO, J.C., OLEKSYK, T.K., BROOKS, L.D., FELSENFELD, A.L., MCEWEN, J.E., CLEMM, N.C., DUNCANSON, A., DUNN, M., GREEN, E.D., GUYER, M.S., PETERSON, J.L., ABECASIS, G.R., AUTON, A., BROOKS, L.D., DEPRISTO, M.A., DURBIN, R.M., HANDSAKER, R.E., KANG, H.M., MARTH, G.T., MCVEAN, G.A. and 1000 Genomes Project Consortium, 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature*, NOV 1 2012, vol. 491, no. 7422, pp. 56-65 ISSN 0028-0836. DOI 10.1038/nature11632.

ALTSHULER, D.M., GIBBS, R.A., PELTONEN, L., DERMITZAKIS, E., SCHAFFNER, S.F., YU, F., BONNEN, P.E., DE BAKKER, P.I.W., DELOUKAS, P., GABRIEL, S.B., GWILLIAM, R., HUNT, S., INOUYE, M., JIA, X., PALOTIE, A., PARKIN, M., WHITTAKER, P., CHANG, K., HAWES, A., LEWIS, L.R., REN, Y., WHEELER, D., MUZNY, D.M., BARNES, C., DARVISHI, K., HURLES, M., KORN, J.M., KRISTIANSSON, K., LEE, C., MCCARROLL, S.A., NEMESH, J., KEINAN, A., MONTGOMERY, S.B., POLLACK, S., PRICE, A.L., SORANZO, N., GONZAGA-JAUREGUI, C., ANTTILA, V., BRODEUR, W., DALY, M.J., LESLIE, S., MCVEAN, G., MOUTSIANAS, L., NGUYEN, H., ZHANG, Q., GHORI, M.J.R., MCGINNIS, R., MCLAREN, W., TAKEUCHI, F., GROSSMAN, S.R., SHLYAKHTER, I., HOSTETTER, E.B., SABETI, P.C., ADEBAMOWO, C.A., FOSTER, M.W., GORDON, D.R., LICINIO, J., MANCA, M.C., MARSHALL, P.A., MATSUDA, I., NGARE, D., WANG, V.O., REDDY, D., ROTIMI, C.N., ROYAL, C.D., SHARP, R.R., ZENG, C., BROOKS, L.D., MCEWEN, J.E. and Int HapMap 3 Consortium, 2010. Integrating common and rare genetic variation in diverse human populations. *Nature*, SEP 2 2010, vol. 467, no. 7311, pp. 52-58 ISSN 0028-0836. DOI 10.1038/nature09298.

AMIGO, J., PHILLIPS, C., LAREU, M. and CARRACEDO, A., 2008a. The SNPforID browser: an online tool for query and display of frequency data from the SNPforID project. *International Journal of Legal Medicine*, SEP 2008, vol. 122, no. 5, pp. 435-440 ISSN 0937-9827. DOI 10.1007/s00414-008-0233-7.

AMIGO, J., SALAS, A., PHILLIPS, C.and CARRACEDO, A., 2008b. *SPSmart: adapting population based SNP genotype databases for fast and comprehensive web access.* ISBN 1471-2105.

Applied Biosystems., 2012. *DNA Fragment Analysis by Capillary Electrophoresis User Guide.* Life Technologies ed., Foster City, USA: Life Technologies Corporation.

Applied Biosystems., 2010a. *Custom TaqMan® SNP Genotyping Assays: Protocol.* Foster City, CA, USA: Applied Biosystems.

Applied Biosystems., 2010b. *Using the SNaPshot® Multiplex System with the POP-7™ Polymer on Applied Biosystems 3730/3730xl DNA Analyzers and 3130/3130xl Genetic Analyzers.* Foster City, CA USA: Life Technologies.

Applied Biosystems., 2009a. *ABI PRISM® SNaPshot™ Multiplex Kit: Protocol.* Foster City, CA: Life Technologies.

Applied Biosystems., 2009b. *Applied Biosystems 3500/3500xLGenetic Analyzer User Guide.* Foster City, CA: Life Technologies.

Applied Biosystems., 2006. *Product bulletin TaqMan® SNP Genotyping Assays.* Foster City, CA, USA: Applied Biosystems.

Applied Biosystems., 2005. *Getting Started Guide, Allelic Discrimination Assays.* Foster City, CA, USA: .

Applied Biosystems., 2004. *Applied Biosystems 3130 and 3130xlGenetic Analyzers: One Polymer, One Capillary Array Length, Maximum Performance.* Foster City, CA: Life Technologies.

BALLANTYNE, K.N., VAN OORSCHOT, R.A.H. and MITCHELL, R.J., 2011. Increased amplification success from forensic samples with locked nucleic acids. *Forensic Science International: Genetics*, 8, vol. 5, no. 4, pp. 276-280 ISSN 1872-4973. DOI 10.1016/j.fsigen.2010.04.001.

BALLANTYNE, K.N., VAN OVEN, M., RALF, A., STONEKING, M., MITCHELL, R.J., VAN OORSCHOT, R.A.H. and KAYSER, M., 2012. MtDNA SNP multiplexes for efficient inference of matrilineal genetic ancestry within Oceania. *Forensic Science International: Genetics*, 7, vol. 6, no. 4, pp. 425-436 ISSN 1872-4973. DOI http://dx.doi.org/10.1016/j.fsigen.2011.08.010.

BALLARD, D.J., PHILLIPS, C., THACKER, C.R., ROBSON, C., REVOIR, A.P. and SYNDERCOMBE COURT, D., 2005. Y chromosome STR haplotypes in three UK populations. *Forensic Science International*, 9/10, vol. 152, no. 2–3, pp. 289-305 ISSN 0379-0738. DOI http://dx.doi.org/10.1016/j.forsciint.2004.10.006.

BARBARO, A., PHILLIPS, C., FONDEVILA, M., LAREU, M. and CARRACEDO, A., 2012. Genetic variability of the SNPforID 52-plex identification SNP panel in Italian population samples.

Forensic Science International-Genetics, DEC 2012, vol. 6, no. 6, pp. E185-E186 ISSN 1872-4973. DOI 10.1016/j.fsigen.2012.07.002.

BEHR, S., MÄTZIG, M., LEVIN, A., EICKHOFF, H. and HELLER, C., 1999. A fully automated multicapillary electrophoresis device for DNA analysis. *Electrophoresis*, vol. 20, no. 7, pp. 1492-1507 ISSN 1522-2683. DOI 10.1002/(SICI)1522-2683(19990601)20:7<1492::AID-ELPS1492>3.0.CO;2-V.

BEKAERT, B., WENS, J. and DECORTE, R., 2011. Optimization and validation of the SNPforID 34-SNPplex for POP7™. *Forensic Science International: Genetics Supplement Series*, 12, vol. 3, no. 1, pp. e43-e44 ISSN 1875-1768. DOI 10.1016/j.fsigss.2011.08.021.

BØRSTING, C., HJORT, B.B. and MORLING, N., 2011. SNP typing of crime case samples with the SNPforID multiplex assay. *Forensic Science International: Genetics Supplement Series*, 12, vol. 3, no. 1, pp. e99-e100 ISSN 1875-1768. DOI http://dx.doi.org/10.1016/j.fsigss.2011.08.049.

BORSTING, C., MIKKELSEN, M. and MORLING, N., 2012a. Kinship Analysis with Diallelic SNPs - Experiences with the SNPforID Multiplex in an ISO17025 Accreditated Laboratory. *Transfusion Medicine and Hemotherapy*, JUN 2012, vol. 39, no. 3, pp. 195-201 ISSN 1660-3796. DOI 10.1159/000338957.

BØRSTING, C., MOGENSEN, H.S. and MORLING, N., 2013. Forensic genetic SNP typing of low-template DNA and highly degraded DNA from crime case samples. *Forensic Science International: Genetics*, 5, vol. 7, no. 3, pp. 345-352 ISSN 1872-4973. DOI 10.1016/j.fsigen.2013.02.004.

BORSTING, C. and MORLING, N., 2011. Mutations and/or close relatives? Six case work examples where 49 autosomal SNPs were used as supplementary markers. *Forensic Science International-Genetics*, JUN 2011, vol. 5, no. 3, pp. 236-241 ISSN 1872-4973. DOI 10.1016/j.fsigen.2010.02.007.

BORSTING, C., ROCKENBAUER, E. and MORLING, N., 2009. Validation of a single nucleotide polymorphism (SNP) typing assay with 49 SNPs for forensic genetic testing in a laboratory accredited according to the ISO 17025 standard. *Forensic Science International-Genetics*, DEC 2009, vol. 4, no. 1, pp. 34-42 ISSN 1872-4973. DOI 10.1016/j.fsigen.2009.04.004.

BORSTING, C., TOMAS, C. and MORLING, N., 2012b. Typing of 49 Autosomal SNPs by Single Base Extension and Capillary Electrophoresis for Forensic Genetic Testing. *Dna Electrophoresis Protocols for Forensic Genetics*, 2012, vol. 830, pp. 87-107 ISSN 1064-3745; 978-1-61779-460-5. DOI 10.1007/978-1-61779-461-2_7.

BOTSTEIN, D., WHITE, R.L., SKOLNICK, M. and DAVIS, R.W., 1980. Construction of a Genetic-Linkage Map in Man using Restriction Fragment Length Polymorphisms. *American Journal of Human Genetics*, 1980, vol. 32, no. 3, pp. 314-331 ISSN 0002-9297.

BRENNER, C.H., 2006. Some mathematical problems in the DNA identification of victims in the 2004 tsunami and similar mass fatalities. *Forensic Science International*, 3/10, vol. 157, no. 2–3, pp. 172-180 ISSN 0379-0738. DOI http://dx.doi.org/10.1016/j.forsciint.2005.11.003.

BRIÓN, M., SANCHEZ, J., BALOGH, K., THACKER, C., BLANCO-VEREA, A., BØRSTING, C., STRADMANN-BELLINGHAUSEN, B., BOGUS, M., SYNDERCOMBE-COURT, D., SCHNEIDER, P., CARRACEDO, A. and MORLING, N., 2005. Introduction of an single nucleodite polymorphism-based ?Major Y-chromosome haplogroup typing kit? suitable for predicting the geographical origin of male lineages. *Electrophoresis*, vol. 26, no. 23, pp. 4411-4420 ISSN 1522-2683. DOI 10.1002/elps.200500293.

BUDOWLE, B. and VAN DAAL, A., 2008. Forensically relevant SNP classes. *BioTechniques*, APR 2008, vol. 44, no. 5, pp. 603-+ ISSN 0736-6205. DOI 10.2144/000112806.

BULBUL, O., PHILLIPS, C., ARGAC, D., SHAHZAD, M.S., FONDEVILLA, M., ACAR, E., ARADAS, A.F., FILOGLU, G. and ALTUNCUL, H., 2009. Internal validation of 29 autosomal SNP multiplex using a ABI 310 genetic analyser. *Forensic Science International: Genetics Supplement Series*, 12, vol. 2, no. 1, pp. 129-130 ISSN 1875-1768. DOI 10.1016/j.fsigss.2009.08.193.

BUTLER, J. M., 2011. *Advanced topics in Forensic DNA typing: Methodology.* Elsevier.

BUTLER, J.M., 2005. *Forensic DNA Typing: Biology, Technology and Genetics for STR Markers.* Second Edition ed. Burlington,MA, USA: Elsevier Academic Press ISBN 0-12-147952-8.

BUTLER, J.M., BUDOWLE, B., GILL, P., KIDD, K.K., PHILLIPS, C., SCHNEIDER, P.M., VALLONE, P.M. and MORLING, N., 2008. Report on ISFG SNP Panel Discussion. *Forensic Science International: Genetics Supplement Series*, 8, vol. 1, no. 1, pp. 471-472 ISSN 1875-1768. DOI http://dx.doi.org/10.1016/j.fsigss.2007.10.159.

BUTLER, J.M., BUEL, E., CRIVELLENTE, F. and MCCORD, B.R., 2004. Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. *Electrophoresis*, JUN 2004, vol. 25, no. 10-11, pp. 1397-1412 ISSN 0173-0835. DOI 10.1002/elps.200305822.

CANN, H.M., DE TOMA, C., CAZES, L., LEGRAND, M.F., MOREL, V., PIOUFFRE, L., BODMER, J., BODMER, W.F., BONNE-TAMIR, B., CAMBON-THOMSEN, A., CHEN, Z., CHU, J.Y., CARCASSI, C., CONTU, L., DU, R.F., EXCOFFIER, L., FERRARA, G.B., FRIEDLAENDER, J.S., GROOT, H., GURWITZ, D., JENKINS, T., HERRERA, R.J., HUANG, X.Y., KIDD, J., KIDD, K.K., LANGANEY, A., LIN, A.A., MEHDI, S.Q., PARHAM, P., PIAZZA, A., PISTILLO, M.P., QIAN, Y.P., SHU, Q.F., XU, J.J., ZHU, S., WEBER, J.L., GREELY, H.T., FELDMAN, M.W., THOMAS, G., DAUSSET, J. and CAVALLI-SFORZA, L.L., 2002. A human genome diversity cell line panel. *Science*, APR 12 2002, vol. 296, no. 5566, pp. 261-262 ISSN 0036-8075.

CAVALLI-SFORZA, L.L., 2005. Opinion - The Human Genome Diversity Project: past, present and future. *Nature Reviews Genetics*, APR 2005, vol. 6, no. 4, pp. 333-340 ISSN 1471-0056. DOI 10.1038/nrg1579.

CHAKRABORTY, R., STIVERS, D.N., SU, B., ZHONG, Y.X. and BUDOWLE, B., 1999. The utility of short tandem repeat loci beyond human identification: Implications for development of new DNA typing systems. *Electrophoresis*, JUN 1999, vol. 20, no. 8, pp. 1682-1696 ISSN 0173-0835. DOI 10.1002/(SICI)1522-2683(19990101)20:8<1682::AID-ELPS1682>3.3.CO;2-Q.

CHANG, Y.M., BOTH, K. and BURGOYNE, L.A., 2002. Genetic variation at three DYS-STR loci and amelogenin between the Australian Caucasians and three Asian ethnic groups. *Journal of Forensic Sciences*, JUL 2002, vol. 47, no. 4, pp. 918-921 ISSN 0022-1198.

CHANG, Y.M., PERUMAL, R., KEAT, P.Y. and 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*, 3/22, vol. 167, no. 1, pp. 70-76 ISSN 0379-0738. DOI 10.1016/j.forsciint.2006.01.002.

CHANG, Y.M., SWARAN, Y., PHOON, Y.K., SOTHIRASAN, K., SIM, H.T., LIM, K.B. and 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*, 6, vol. 3, no. 3, pp. e77-e80 ISSN 1872-4973. DOI http://dx.doi.org/10.1016/j.fsigen.2008.07.007.

COBLE, M.D., JUST, R.S., O'CALLAGHAN, J.E., LETMANYI, I.H., PETERSON, C.T., IRWIN, J.A. and PARSONS, T.J., 2004. Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians. *International Journal of Legal Medicine*, JUN 2004, vol. 118, no. 3, pp. 137-146 ISSN 0937-9827. DOI 10.1007/s00414-004-0427-6.

COBLE, M.D., LOREILLE, O.M., WADHAMS, M.J., EDSON, S.M., MAYNARD, K., MEYER, C.E., NIEDERSTAETTER, H., BERGER, C., BERGER, B., FALSETTI, A.B., GILL, P., PARSON, W. and FINELLI, L.N., 2009. Mystery Solved: The Identification of the Two Missing Romanov Children Using DNA Analysis. *Plos One*, MAR 11 2009, vol. 4, no. 3, pp. e4838 ISSN 1932-6203. DOI 10.1371/journal.pone.0004838.

CORRADINI, D. and SPRECACENERE, L., 2003. Dependence of the electroosmotic flow in bare fused-silica capillaries from pH, ionic strength and composition of electrolyte solutions tailored for protein capillary zone electrophoresis. *Chromatographia*, NOV 2003, vol. 58, no. 9-10, pp. 587-596 ISSN 0009-5893. DOI 10.1365/s10337-003-0098-3.

DE LA VEGA, F.A., LAZARUK, K.D., RHODES, M.D. and WENZ, M.H., 2005a. Assessment of two flexible and compatible SNP genotyping platforms: TaqMan (R) SNP genotyping assays and the SNPlex (TM) genotyping system. *Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis*, JUN 3 2005, vol. 573, no. 1-2, pp. 111-135 ISSN 0027-5107. DOI 10.1016/j.mrfmmm.2005.01.008.

DE LA VEGA, F.A., LAZARUK, K.D., RHODES, M.D. and WENZ, M.H., 2005b. Assessment of two flexible and compatible SNP genotyping platforms: TaqMan (R) SNP genotyping assays and the SNPlex (TM) genotyping system. *Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis*, JUN 3 2005, vol. 573, no. 1-2, pp. 111-135 ISSN 0027-5107. DOI 10.1016/j.mrfmmm.2005.01.008.

DEPARTMENT OF STATISTICS, M., 2010. *Population distribution and basic demographic characteristics.* Putrajaya, Malaysia: Department of Satistics, Malaysia.

DETWILER, M.M., HAMP, T.J. and KAZIM, A.L., 2004. DNA sequencing using the liquid polymer POP-7 (TM) on an ABI PRISM (R) 3100 genetic analyzer. *BioTechniques*, JUN 2004, vol. 36, no. 6, pp. 932-933 ISSN 0736-6205.

DIXON, L.A., MURRAY, C.M., ARCHER, E.J., DOBBINS, A.E., KOUMI, P. and GILL, P., 2005. Validation of a 21-locus autosomal SNP multiplex for forensic identification purposes. *Forensic Science International*, NOV 10 2005, vol. 154, no. 1, pp. 62-77 ISSN 0379-0738. DOI 10.1016/j.forsciint.2004.12.011.

DROBNIC, K., BORSTING, C., ROCKENBAUER, E., TOMAS, C. and MORLING, N., 2010. Typing of 49 autosomal SNPs by SNaPshot (R) in the Slovenian population. *Forensic Science International-Genetics*, OCT 2010, vol. 4, no. 5, pp. E125-E127 ISSN 1872-4973. DOI 10.1016/j.fsigen.2010.01.014.

DULIK, M., ZHADANOV, S., OSIPOVA, L., ASKAPULI, A., GAU, L., GOKCUMEN, O., RUBINSTEIN, S. and SCHURR, T., 2012. Mitochondrial DNA and Y Chromosome Variation Provides Evidence for a Recent Common Ancestry between Native Americans and Indigenous Altaians. *The American Journal of Human Genetics*, 2/10, vol. 90, no. 2, pp. 229-246 ISSN 0002-9297. DOI http://dx.doi.org/10.1016/j.ajhg.2011.12.014.

EXCOFFIER, L. and LISCHER, H.E.L., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, vol. 10, no. 3, pp. 564-567 ISSN 1755-0998. DOI 10.1111/j.1755-0998.2010.02847.x.

FELHOFER, J.L., BLANES, L. and GARCIA, C.D., 2010. Recent developments in instrumentation for capillary electrophoresis and microchip-capillary electrophoresis. *Electrophoresis*, vol. 31, no. 15, pp. 2469-2486 ISSN 1522-2683. DOI 10.1002/elps.201000203.

FONDEVILA, M., PHILLIPS, C., SANTOS, C., FREIRE ARADAS, A., VALLONE, P.M., BUTLER, J.M., LAREU, M.V. and CARRACEDO, A., 2013a. Revision of the SNPforID 34-plex forensic ancestry test: Assay enhancements, standard reference sample genotypes and extended population studies. *Forensic Science International-Genetics*, JAN 2013, vol. 7, no. 1, pp. 63-74 ISSN 1872-4973. DOI 10.1016/j.fsigen.2012.06.007.

FONDEVILA, M., PHILLIPS, C., SANTOS, C., FREIRE ARADAS, A., VALLONE, P.M., BUTLER, J.M., LAREU, M.V. and CARRACEDO, Á, 2013b. Revision of the SNPforID 34-plex forensic ancestry test: Assay enhancements, standard reference sample genotypes and extended population studies. *Forensic Science International: Genetics*, 1, vol. 7, no. 1, pp. 63-74 ISSN 1872-4973. DOI 10.1016/j.fsigen.2012.06.007.

FRAZER, K.A., BALLINGER, D.G., COX, D.R., HINDS, D.A., STUVE, L.L., GIBBS, R.A., BELMONT, J.W., BOUDREAU, A., HARDENBOL, P., LEAL, S.M., PASTERNAK, S., WHEELER, D.A., WILLIS, T.D., YU, F., YANG, H., ZENG, C., GAO, Y., HU, H., HU, W., LI, C., LIN, W., LIU, S., PAN, H., TANG, X., WANG, J., WANG, W., YU, J., ZHANG, B., ZHANG, Q., ZHAO, H., ZHAO, H., ZHOU, J., GABRIEL, S.B., BARRY, R., BLUMENSTIEL, B., CAMARGO, A., DEFELICE, M., FAGGART, M., GOYETTE, M., GUPTA, S., MOORE, J., NGUYEN, H., ONOFRIO, R.C., PARKIN, M., ROY, J., STAHL, E., WINCHESTER, E., ZIAUGRA, L., ALTSHULER, D., SHEN, Y., YAO, Z., HUANG, W., CHU, X., HE, Y., JIN, L., LIU, Y., SHEN, Y., SUN, W., WANG, H., WANG, Y., WANG, Y., XIONG, X., XU, L., WAYE, M.M.Y., TSUI, S.K.W., WONG, J.T., GALVER, L.M., FAN, J., GUNDERSON, K., MURRAY, S.S., OLIPHANT, A.R., CHEE, M.S., MONTPETIT, A., CHAGNON, F., FERRETTI, V., LEBOEUF, M., OLIVIER, J., PHILLIPS, M.S., ROUMY, S., SALLEE, C., VERNER, A., HUDSON, T.J., KWOK, P., CAI, D., KOBOLDT, D.C., MILLER, R.D., PAWLIKOWSKA, L., TAILLON-MILLER, P., XIAO, M., TSUI, L., MAK, W., SONG, Y.Q., TAM, P.K.H., NAKAMURA, Y., KAWAGUCHI, T., KITAMOTO, T., MORIZONO, T., NAGASHIMA, A., OHNISHI, Y., SEKINE, A., TANAKA, T., TSUNODA, T., DELOUKAS, P., BIRD, C.P., DELGADO, M.,

DERMITZAKIS, E.T., GWILLIAM, R., HUNT, S., MORRISON, J., POWELL, D., STRANGER, B.E., WHITTAKER, P., BENTLEY, D.R., DALY, M.J., DE BAKKER, P.I.W., BARRETT, J., CHRETIEN, Y.R., MALLER, J., MCCARROLL, S., PATTERSON, N., PE'ER, I., PRICE, A., PURCELL, S., RICHTER, D.J., SABETI, P., SAXENA, R., SCHAFFNER, S.F., SHAM, P.C., VARILLY, P., ALTSHULER, D., STEIN, L.D., KRISHNAN, L., SMITH, A.V., TELLO-RUIZ, M.K., THORISSON, G.A., CHAKRAVARTI, A., CHEN, P.E., CUTLER, D.J., KASHUK, C.S., LIN, S., ABECASIS, G.R., GUAN, W., LI, Y., MUNRO, H.M., QIN, Z.S., THOMAS, D.J., MCVEAN, G., AUTON, A., BOTTOLO, L., CARDIN, N., EYHERAMENDY, S., FREEMAN, C., MARCHINI, J., MYERS, S., SPENCER, C., STEPHENS, M., DONNELLY, P., CARDON, L.R., CLARKE, G., EVANS, D.M., MORRIS, A.P., WEIR, B.S., TSUNODA, T., JOHNSON, T.A., MULLIKIN, J.C., SHERRY, S.T., FEOLO, M., SKOL, A. and Int HapMap Consortium, 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature*, OCT 18 2007, vol. 449, no. 7164, pp. 851-U3 ISSN 0028-0836. DOI 10.1038/nature06258.

FREDLAKE, C.P., HERT, D.G., ROOT, B.E. and BARRON, A.E., 2008. Polymer systems designed specifically for DNA sequencing by microchip electrophoresis: A comparison with commercially available materials. *Electrophoresis*, vol. 29, no. 23, pp. 4652-4662 ISSN 1522-2683. DOI 10.1002/elps.200800352.

GE, J., BUDOWLE, B., PLANZ, J.V. and CHAKRABORTY, R., 2010. Haplotype block: a new type of forensic DNA markers. *International Journal of Legal Medicine*, SEP 2010, vol. 124, no. 5, pp. 353-361 ISSN 0937-9827. DOI 10.1007/s00414-009-0400-5.

GIBBONS, A., 2011. Aboriginal Genome Shows Two-Wave Settlement of Asia. *Science*, September 23, vol. 333, no. 6050, pp. 1689-1691 DOI 10.1126/science.333.6050.1689.

GIBSON, N.J., 2006. The use of real-time PCR methods in DNA sequence variation analysis. *Clinica Chimica Acta*, 1, vol. 363, no. 1–2, pp. 32-47 ISSN 0009-8981. DOI 10.1016/j.cccn.2005.06.022.

GILL, P., 2001. An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. *International Journal of Legal Medicine*, APR 2001, vol. 114, no. 4-5, pp. 204-210 ISSN 0937-9827. DOI 10.1007/s004149900117.

GILL, P., WERRETT, D.J., BUDOWLE, B. and GUERRIERI, R., 2004. An assessment of whether SNPs will replace STRs in national DNA databases. *Science & Justice*, JAN-MAR 2004, vol. 44, no. 1, pp. 51-53 ISSN 1355-0306. DOI 10.1016/S1355-0306(04)71685-8.

GILL, P. and BUCKLETON, J., 2009. Low copy number typing—Where next?. *Forensic Science International: Genetics Supplement Series*, 12, vol. 2, no. 1, pp. 553-555 ISSN 1875-1768. DOI 10.1016/j.fsigss.2009.08.153.

GOLENBERG, E.M., BICKEL, A. and WEIHS, P., 1996. Effect of Highly Fragmented DNA on PCR. *Nucleic Acids Research*, December 01, vol. 24, no. 24, pp. 5026-5033 DOI 10.1093/nar/24.24.5026.

Goodwin, W., Linacre, A., Hadi, Sibte., 2011. *An introduction to Forensic Genetics.* Second Edition ed.Wiley-Blackwell.

GRIMES, E.A., NOAKE, P.J., DIXON, L. and URQUHART, A., 2001. Sequence polymorphism in the human melanocortin 1 receptor gene as an indicator of the red hair phenotype. *Forensic*

Science International, 11/1, vol. 122, no. 2–3, pp. 124-129 ISSN 0379-0738. DOI http://dx.doi.org/10.1016/S0379-0738(01)00480-7.

GUSMAO, L., ALVES, C., GOMES, I. and SANCHEZ-DIZ, P., 2012. Capillary Electrophoresis of an X-Chromosome STR Decaplex for Kinship Deficiency Cases. *DNA Electrophoresis Protocols for Forensic Genetics*, 2012, vol. 830, pp. 57-71 ISSN 1064-3745(print); 978-1-61779-460-5(H). DOI 10.1007/978-1-61779-461-2_5.

HARAYAMA, Y., OKI, T., HAYASHI, T., TSUKADA, K., OTA, M. and ASAMURA, H., 2011. A comparative study on the techniques of SNPs analysis for degraded DNA samples. *Forensic Science International: Genetics Supplement Series*, 12, vol. 3, no. 1, pp. e180-e181 ISSN 1875-1768. DOI http://dx.doi.org/10.1016/j.fsigss.2011.08.090.

HATIN, W.I., NUR-SHAFAWATI, A.R., ZAHRI, M., XU, S., JIN, L., TAN, S., RIZMAN-IDID, M., ZILFALIL, B.A. and HUGO Pan-Asian SNP Consortium, 2011. Population Genetic Structure of Peninsular Malaysia Malay Sub-Ethnic Groups. *Plos One*, APR 5 2011, vol. 6, no. 4, pp. e18312 ISSN 1932-6203. DOI 10.1371/journal.pone.0018312.

HOCHMEISTER, M.N., BUDOWLE, B., BORER, U.V., EGGMANN, U., COMEY, C.T. and DIRNHOFER, R., 1991. Typing of Deoxyribonucleic-Acid (Dna) Extracted from Compact-Bone from Human Remains. *Journal of Forensic Sciences*, NOV 1991, vol. 36, no. 6, pp. 1649-1661 ISSN 0022-1198.

HOU, Y.P., ZHANG, J., LI, Y.B., WU, J., ZHANG, S.Z. and PRINZ, M., 2001. Allele sequences of six new Y-STR loci and haplotypes in the Chinese Han population. *Forensic Science International*, 5/15, vol. 118, no. 2–3, pp. 147-152 ISSN 0379-0738. DOI http://dx.doi.org/10.1016/S0379-0738(01)00383-8.

HUGHES-STAMM, S.R., ASHTON, K.J. and VAN DAAL, A., 2011. Assessment of DNA degradation and the genotyping success of highly degraded samples. *International Journal of Legal Medicine*, MAY 2011, vol. 125, no. 3, pp. 341-348 ISSN 0937-9827. DOI 10.1007/s00414-010-0455-3.

ISENBERG, A.R., ALLEN, R.O., KEYS, K.M., SMERICK, J.B., BUDOWLE, B. and MCCORD, B.R., 1998. Analysis of two multiplexed short tandem repeat systems using capillary electrophoresis with multiwavelength fluorescence detection. *Electrophoresis*, JAN 1998, vol. 19, no. 1, pp. 94-100 ISSN 0173-0835. DOI 10.1002/elps.1150190117.

IZUAN, M., SEAH, L.H., PANNEERCHELVAM, S. and NOR, N.M., 2005. STR data for the PowerPlex (R) 16 system loci for the Malays, Chinese and Indians groups of the Malaysian population. *Journal of Forensic Sciences*, SEP 2005, vol. 50, no. 5, pp. 1225-1228 ISSN 0022-1198.

JAKOA., 2013. *Sub-tribe/Ethnic.* JAKOA, Ministry of Rural and Regional Development, Malaysia.

JEFFREYS, A.J., WILSON, V. and THEIN, S.L., 1985. Hypervariable 'minisatellite' regions in human DNA. *Nature*, vol. 314, no. 6006, pp. 67-73 DOI 10.1038/314067a0.

JOBLING, M.A. and KING, T.E., 2004. The distribution of Y-chromosomal haplotypes: forensic implications. *International Congress Series*, 4, vol. 1261, no. 0, pp. 70-72 ISSN 0531-5131. DOI http://dx.doi.org/10.1016/S0531-5131(03)01787-4.

JOBLING, M, HOLLOX, E., HURLES,M., KIVISILD, T. and TYLER-SMITH, C., 2013. Human evolutionary genetics, 2nd Edition, Garland Science.

John M. Butler, Michael D. Coble, Peter M. Vallone, 2007. STRs vs. SNPs: thoughts on the future of forensic DNA testing, vol. 3, pp. 200 DOI DOI 10.1007/s12024-007-0018-1.

JONES, D.A., 1972. Blood Samples : Probability of Discrimination. *Journal of the Forensic Science Society*, 4, vol. 12, no. 2, pp. 355-359 ISSN 0015-7368. DOI 10.1016/S0015-7368(72)70695-7.

KAYSER, M. and SCHNEIDER, P.M., 2009. DNA-based prediction of human externally visible characteristics in forensics: Motivations, scientific challenges, and ethical considerations. *Forensic Science International: Genetics*, 6, vol. 3, no. 3, pp. 154-161 ISSN 1872-4973. DOI http://dx.doi.org/10.1016/j.fsigen.2009.01.012.

KIDD, K.K., KIDD, J.R., SPEED, W.C., FANG, R., FURTADO, M.R., HYLAND, F.C.L. and PAKSTIS, A.J., 2012. Expanding data and resources for forensic use of SNPs in individual identification. *Forensic Science International-Genetics*, SEP 2012, vol. 6, no. 5, pp. 646-652 ISSN 1872-4973. DOI 10.1016/j.fsigen.2012.02.012.

KIDD, K.K., PAKSTIS, A.J., SPEED, W.C., GRIGORENKO, E.L., KAJUNA, S.L.B., KAROMA, N.J., KUNGULILO, S., KIM, J., LU, R., ODUNSI, A., OKONOFUA, F., PARNAS, J., SCHULZ, L.O., ZHUKOVA, O.V. and KIDD, J.R., 2006. Developing a SNP panel for forensic identification of individuals. *Forensic Science International*, 12/1, vol. 164, no. 1, pp. 20-32 ISSN 0379-0738. DOI 10.1016/j.forsciint.2005.11.017.

KOUMI, P., GREEN, H., HARTLEY, S., JORDAN, D., LAHEC, S., LIVETT, R., TSANG, K. and WARD, D., 2004. Evaluation and validation of the ABI 3700, ABI 3100 and the MegaBACE 1000 capillary array electrophoresis instruments for use with short tandem repeat microsatellite typing in a forensic environment. *Electrophoresis*, JUL, vol. 25, no. 14, pp. 2227-2241 ISSN 0173-0835. DOI 10.1002/elps.200305976.

LAMASON, R.L., MOHIDEEN, M.P.K., MEST, J.R., WONG, A.C., NORTON, H.L., AROS, M.C., JURYNEC, M.J., MAO, X., HUMPHREVILLE, V.R., HUMBERT, J.E., SINHA, S., MOORE, J.L., JAGADEESWARAN, P., ZHAO, W., NING, G., MAKALOWSKA, I., MCKEIGUE, P.M., O'DONNELL, D., KITTLES, R., PARRA, E.J., MANGINI, N.J., GRUNWALD, D.J., SHRIVER, M.D., CANFIELD, V.A. and CHENG, K.C., 2005. SLC24A5, a Putative Cation Exchanger, Affects Pigmentation in Zebrafish and Humans. *Science*, December 16, vol. 310, no. 5755, pp. 1782-1786 DOI 10.1126/science.1116238.

LAO, O., VALLONE, P.M., COBLE, M.D., DIEGOLI, T.M., VAN OVEN, M., VAN DER GAAG, K.J., PIJPE, J., DE KNIJFF, P. and KAYSER, M., 2010. Evaluating self-declared ancestry of U.S. Americans with autosomal, Y-chromosomal and mitochondrial DNA. *Human Mutation*, vol. 31, no. 12, pp. E1875-E1893 ISSN 1098-1004. DOI 10.1002/humu.21366.

LARUE, B.L., GE, J., KING, J.L. and 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*, JUL 2012, vol. 126, no. 4, pp. 533-540 ISSN 0937-9827. DOI 10.1007/s00414-012-0667-9.

LAZARUK, K., WALSH, P.S., OAKS, F., GILBERT, D., ROSENBLUM, B.B., MENCHEN, S., SCHEIBLER, D., WENZ, H.M., HOLT, C. and WALLIN, J., 1998. Genotyping of forensic short tandem repeat (STR) systems based on sizing precision in a capillary electrophoresis instrument. *Electrophoresis*, JAN 1998, vol. 19, no. 1, pp. 86-93 ISSN 0173-0835. DOI 10.1002/elps.1150190116.

LEE, S.B., CLABAUGH, K.C., SILVA, B., ODIGIE, K.O., COBLE, M.D., LOREILLE, O., SCHEIBLE, M., FOURNEY, R.M., STEVENS, J., CARMODY, G.R., PARSONS, T.J., POZDER, A., EISENBERG, A.J., BUDOWLE, B., AHMAD, T., MILLER, R.W. and CROUSE, C.A., 2012. Assessing a novel room temperature DNA storage medium for forensic biological samples. *Forensic Science International-Genetics*, JAN 2012, vol. 6, no. 1, pp. 31-40 ISSN 1872-4973. DOI 10.1016/j.fsigen.2011.01.008.

LI, J.Z., ABSHER, D.M., TANG, H., SOUTHWICK, A.M., CASTO, A.M., RAMACHANDRAN, S., CANN, H.M., BARSH, G.S., FELDMAN, M., CAVALLI-SFORZA, L.L. and MYERS, R.M., 2008. Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation. *Science*, February 22, vol. 319, no. 5866, pp. 1100-1104 DOI 10.1126/science.1153717.

LI, L., LI, C., ZHANG, S., ZHAO, S., LIU, Y. and LIN, Y., 2010. Analysis of 14 highly informative SNP markers on X chromosome by TaqMan (R) SNP genotyping assay. *Forensic Science International-Genetics*, OCT 2010, vol. 4, no. 5, pp. E145-E148 ISSN 1872-4973. DOI 10.1016/j.fsigen.2010.04.004.

Life Technologies., 2013. *Quality value in TaqMan® Genotyper Software.* Life Technologies.

LIM, K.B., JEEVAN, N.H., JAYA, P., OTHMAN, M.I. and LEE, Y.H., 2001. STR data for the AmpFlSTR Profiler loci from the three main ethnic population groups (Malay, Chinese and Indian) in Malaysia. *Forensic Science International*, JUN 1 2001, vol. 119, no. 1, pp. 109-112 ISSN 0379-0738. DOI 10.1016/S0379-0738(00)00386-8.

MADABHUSHI, R., 2001. DNA sequencing in noncovalently coated capillaries using low viscosity polymer solutions. *Methods in Molecular Biology (Clifton, N.J.)*, 2001, vol. 163, pp. 309-15 ISSN 1064-3745.

MADABHUSHI, R.S., 1998. Separation of 4-color DNA sequencing extension products in noncovalently coated capillaries using low viscosity polymer solutions. *Electrophoresis*, FEB 1998, vol. 19, no. 2, pp. 224-230 ISSN 0173-0835.

MADABHUSHI, R.S., VAINER, M., DOLNIK, V., ENAD, S., BARKER, D.L., HARRIS, D.W. and MANSFIELD, E.S., 1997. Versatile low-viscosity sieving matrices for nondenaturing DNA separations using capillary array electrophoresis. *Electrophoresis*, vol. 18, no. 1, pp. 104-111 ISSN 1522-2683. DOI 10.1002/elps.1150180120.

MANDABHUSHI, R.S., 1998. Separation of 4-color DNA sequencing extension products in noncovalently coated capillaries using low viscosity polymer solutions. *Electrophoresis*, vol. 19, no. 2, pp. 224-230 ISSN 1522-2683. DOI 10.1002/elps.1150190215.

MARUYAMA, K., 2005. *Genome sequencing technology: improvement of the electrophoretic sequencing process and analysis of the sequencing tool industry.* Leaders for Manufacturing Program. ed., Thesis (M.B.A.) ed. Massachusetts Institute of Technology. Dept. of Chemical

Engineering.; Sloan School of Management.; Leaders for Manufacturing Program.: Massachusetts Institute of Technology.

MIZUNO, N., KITAYAMA, T., FUJII, K., NAKAHARA, H., YOSHIDA, K., SEKIGUCHI, K., YONEZAWA, N., NAKANO, M. and KASAI, K., 2010. A forensic method for the simultaneous analysis of biallelic markers identifying Y chromosome haplogroups inferred as having originated in Asia and the Japanese archipelago. *Forensic Science International-Genetics*, FEB 2010, vol. 4, no. 2, pp. 73-79 ISSN 1872-4973. DOI 10.1016/j.fsigen.2009.06.001.

MORETTI, T.R., BAUMSTARK, A.L., DEFENBAUGH, D.A., KEYS, K.M., BROWN, A.L. and BUDOWLE, B., 2001. Validation of STR typing by capillary electrophoresis. *Journal of Forensic Sciences*, MAY 2001, vol. 46, no. 3, pp. 661-676 ISSN 0022-1198.

MUSGRAVE-BROWN, E., BALLARD, D., BALOGH, K., BENDER, K., BERGER, B., BOGUS, M., BORSTING, C., BRION, M., FORLDEVILA, M., HARRISON, C., OGUZTURUN, C., PARSON, W., PHILLIPS, C., PROFF, C., RAMOS-LUIS, E., SANCHEZ, J.J., SANCHEZ DIZ, P., SOBRINO REY, B., STRADMANN-BELLINGHAUSEN, B., THACKER, C., CARRACEDO, A., MORLING, N., SCHEITHAUER, R., SCHNEIDER, P.M. and COURT, D.S., 2007. Forensic validation of the SNPforID 52-plex assay. *Forensic Science International-Genetics*, JUN 2007, vol. 1, no. 2, pp. 186-190 ISSN 1872-4973. DOI 10.1016/j.fsigen.2007.01.004.

NICODEME, P., 1997. *Selecting optimal oligonucleotide primers for multiplex PCR. Proc Int Conf Intell Syst Mol Biol. 1997;5:210–213.* J.M. STEYAERT ed., 5th ed. Proc Int Conf Intell Syst Mol Biol.

NORMILE, D., 2009. SNP Study Supports Southern Migration Route to Asia. *Science*, December 11, vol. 326, no. 5959, pp. 1470-1470 DOI 10.1126/science.326.5959.1470.

PAKSTIS, A.J., SPEED, W.C., FANG, R., HYLAND, F.C.L., FURTADO, M.R., KIDD, J.R. and KIDD, K.K., 2010. SNPs for a universal individual identification panel. *Human Genetics*, MAR 2010, vol. 127, no. 3, pp. 315-324 ISSN 0340-6717. DOI 10.1007/s00439-009-0771-1.

PAKSTIS, A.J., SPEED, W.C., KIDD, J.R. and KIDD, K.K., 2007. Candidate SNPs for a universal individual identification panel. *Human Genetics*, MAY 2007, vol. 121, no. 3-4, pp. 305-317 ISSN 0340-6717. DOI 10.1007/s00439-007-0342-2.

PEMBERTON, T.J., DEGIORGIO, M. and ROSENBERG, N.A., 2013. Population Structure in a Comprehensive Genomic Data Set on Human Microsatellite Variation. *G3-Genes Genomes Genetics*, MAY 1 2013, vol. 3, no. 5, pp. 891-907 ISSN 2160-1836. DOI 10.1534/g3.113.005728.

PEREIRA, R., FONDEVILA, M., PHILLIPS, C., AMORIM, A., CARRACEDO, A. and GUSMÃO, L., 2008. Genetic characterization of 52 autosomal SNPs in two sub-Saharan African populations. *Forensic Science International: Genetics Supplement Series*, 8, vol. 1, no. 1, pp. 361-363 ISSN 1875-1768. DOI http://dx.doi.org/10.1016/j.fsigss.2007.10.155.

PEREIRA, R. and 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*, 2012, vol. 830, pp. 141-157 ISSN 1064-3745(print); 978-1-61779-460-5(H). DOI 10.1007/978-1-61779-461-2_10.

PEREIRA, R., PHILLIPS, C., PINTO, N., SANTOS, C., BATISTA DOS SANTOS, S.E., AMORIM, A., CARRACEDO, A. and GUSMAO, L., 2012. Straightforward Inference of Ancestry and Admixture Proportions through Ancestry-Informative Insertion Deletion Multiplexing. *Plos One*, JAN 17 2012, vol. 7, no. 1, pp. e29684 ISSN 1932-6203. DOI 10.1371/journal.pone.0029684.

PFAFF, C.L., BARNHOLTZ-SLOAN, J., WAGNER, J.K. and LONG, J.C., 2004. Information on ancestry from genetic markers. *Genetic Epidemiology*, vol. 26, no. 4, pp. 305-315 ISSN 1098-2272. DOI 10.1002/gepi.10319.

PHILLIPS, C., FANG, R., BALLARD, D., FONDEVILA, M., HARRISON, C., HYLAND, F., MUSGRAVE-BROWN, E., PROFF, C., RAMOS-LUIS, E., SOBRINO, B., CARRACEDO, A., FURTADO, M.R., COURT, D.S., SCHNEIDER, P.M. and SNPforID Consortium, 2007a. Evaluation of the Genplex SNP typing system and a 49plex forensic marker panel. *Forensic Science International-Genetics*, JUN 2007, vol. 1, no. 2, pp. 180-185 ISSN 1872-4973. DOI 10.1016/j.fsigen.2007.02.007.

PHILLIPS, C., FONDEVILA, M., GARCÍA-MAGARIÑOS, M., RODRIGUEZ, A., SALAS, A., CARRACEDO, Á and LAREU, M.V., 2008a. Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers. *Forensic Science International: Genetics*, 6, vol. 2, no. 3, pp. 198-204 ISSN 1872-4973. DOI 10.1016/j.fsigen.2008.02.002.

PHILLIPS, C., FONDEVILA, M., GARCÍA-MAGARIÑOS, M., RODRIGUEZ, A., SALAS, A., CARRACEDO, Á and LAREU, M.V., 2008b. Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers. *Forensic Science International: Genetics*, 6, vol. 2, no. 3, pp. 198-204 ISSN 1872-4973. DOI 10.1016/j.fsigen.2008.02.002.

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. and CARRACEDO, Á, 2007b. Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. *Forensic Science International: Genetics*, 12, vol. 1, no. 3–4, pp. 273-280 ISSN 1872-4973. DOI http://dx.doi.org/10.1016/j.fsigen.2007.06.008.

PHILLIPS, C., LAREU, V., SALAS, A. and CARRACEDO, A., 2004. Nonbinary single-nucleotide polymorphism markers. *International Congress Series*, 4, vol. 1261, no. 0, pp. 27-29 ISSN 0531-5131. DOI http://dx.doi.org/10.1016/j.ics.2003.12.008.

PORRAS, L., PHILLIPS, C., FONDEVILA, M., BELTRAN, L., ORTIZ, T., RONDON, F., BARRETO, G., LAREU, M.V., HENAO, J. and CARRACEDO, A., 2009. Genetic variability of the SNPforID 52-plex identification-SNP panel in Central West Colombia. *Forensic Science International-Genetics*, DEC 2009, vol. 4, no. 1, pp. E9-E10 ISSN 1872-4973. DOI 10.1016/j.fsigen.2008.12.003.

POULSEN, L., BORSTING, C., TOMAS, C., GONZALEZ-ANDRADE, F., LOPEZ-PULLES, R., GONZALEZ-SOLORZANO, J. and MORLING, N., 2011. Typing of Amerindian Kichwas and Mestizos from Ecuador with the SNPforID multiplex. *Forensic Science International-Genetics*, AUG 2011, vol. 5, no. 4, pp. E105-E107 ISSN 1872-4973. DOI 10.1016/j.fsigen.2011.03.006.

QIAGEN., 2010a. *QIAamp DNA Investigator Handbook.*

QIAGEN., 2010b. *QIAamp DNA Micro Kit Handbook.* Second edition ed.

RACHLIN, J., DING, C., CANTOR, C. and KASIF, S., 2005a. Computational tradeoffs in multiplex PCR assay design for SNP genotyping. *BMC Genomics*, vol. 6, no. 1, pp. 102 ISSN 1471-2164.

RACHLIN, J., DING, C., CANTOR, C. and KASIF, S., 2005b. MuPlex: multi-objective multiplex PCR assay design. *Nucleic Acids Research*, July 01, vol. 33, no. suppl 2, pp. W544-W547 DOI 10.1093/nar/gki377.

RAGOUSSIS, J., 2006. Genotyping technologies for all. *Drug Discovery Today: Technologies*, 0, vol. 3, no. 2, pp. 115-122 ISSN 1740-6749. DOI 10.1016/j.ddtec.2006.06.013.

RASMUSSEN, M., GUO, X., WANG, Y., LOHMUELLER, K.E., RASMUSSEN, S., ALBRECHTSEN, A., SKOTTE, L., LINDGREEN, S., METSPALU, M., JOMBART, T., KIVISILD, T., ZHAI, W., ERIKSSON, A., MANICA, A., ORLANDO, L., DE LA VEGA, F.M., TRIDICO, S., METSPALU, E., NIELSEN, K., ÁVILA-ARCOS, M.C., MORENO-MAYAR, J.V., MULLER, C., DORTCH, J., GILBERT, M.T.P., LUND, O., WESOLOWSKA, A., KARMIN, M., WEINERT, L.A., WANG, B., LI, J., TAI, S., XIAO, F., HANIHARA, T., VAN DRIEM, G., JHA, A.R., RICAUT, F., DE KNIJFF, P., MIGLIANO, A.B., GALLEGO ROMERO, I., KRISTIANSEN, K., LAMBERT, D.M., BRUNAK, S., FORSTER, P., BRINKMANN, B., NEHLICH, O., BUNCE, M., RICHARDS, M., GUPTA, R., BUSTAMANTE, C.D., KROGH, A., FOLEY, R.A., LAHR, M.M., BALLOUX, F., SICHERITZ-PONTÉN, T., VILLEMS, R., NIELSEN, R., WANG, J. and WILLERSLEV, E., 2011. An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia. *Science*, October 07, vol. 334, no. 6052, pp. 94-98 DOI 10.1126/science.1211177.

ROSENBERG, N.A., PRITCHARD, J.K., WEBER, J.L., CANN, H.M., KIDD, K.K., ZHIVOTOVSKY, L.A.and FELDMAN, M.W., 2002. *Genetic Structure of Human Populations.* , December 20,.

ROSENBLUM, B.B., OAKS, F., MENCHEN, S. and JOHNSON, B., 1997. Improved single-strand DNA sizing accuracy in capillary electrophoresis. *Nucleic Acids Research*, OCT 1 1997, vol. 25, no. 19, pp. 3925-3929 ISSN 0305-1048. DOI 10.1093/nar/25.19.3925.

RUIZ, Y., CHIURILLO, M.A., BORJAS, L., PHILLIPS, C., LAREU, M.V. and CARRACEDO, A., 2012. Analysis of the SNPforID 52-plex markers in four Native American populations from Venezuela. *Forensic Science International-Genetics*, SEP 2012, vol. 6, no. 5, pp. E142-E145 ISSN 1872-4973. DOI 10.1016/j.fsigen.2012.02.007.

SABETI, P.C., VARILLY, P., FRY, B., LOHMUELLER, J., HOSTETTER, E., COTSAPAS, C., XIE, X., BYRNE, E.H., MCCARROLL, S.A., GAUDET, R., SCHAFFNER, S.F., LANDER, E.S. and Int HapMap Consortium, 2007. Genome-wide detection and characterization of positive selection in human populations. *Nature*, OCT 18 2007, vol. 449, no. 7164, pp. 913-U12 ISSN 0028-0836. DOI 10.1038/nature06250.

SACHIDANANDAM, R., WEISSMAN, D., SCHMIDT, S.C., KAKOL, J.M., STEIN, L.D., MARTH, G., SHERRY, S., MULLIKIN, J.C., MORTIMORE, B.J., WILLEY, D.L., HUNT, S.E., COLE, C.G., COGGILL, P.C., RICE, C.M., NING, Z., ROGERS, J., BENTLEY, D.R., KWOK, P.Y., MARDIS, E.R., YEH, R.T., SCHULTZ, B., COOK, L., DAVENPORT, R., DANTE, M., FULTON, L., HILLIER, L., WATERSTON, R.H., MCPHERSON, J.D., GILMAN, B., SCHAFFNER, S., VAN ETTEN, W., REICH, D., HIGGINS, J., DALY, M.J., BLUMENSTIEL, B., BALDWIN, J., STANGE-THOMANN, N., ZODY, M.C., LINTON, L., LANDER, E.S. and ALTSHULER, D., 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, vol. 409, no. 6822, pp. 928-933.

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. and MORLING, N., 2006a. A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis*, MAY 2006, vol. 27, no. 9, pp. 1713-1724 ISSN 0173-0835. DOI 10.1002/elps.200500671.

SANCHEZ, J.J., PHILLIPS, C., BØRSTING, C., BOGUS, M., CARRACEDO, A., SYNDERCOMBE-COURT, D., FONDEVILA, M., HARRISON, C.D., MORLING, N., BALOGH, K. and SCHNEIDER, P.M., 2006b. Development of a multiplex PCR assay detecting 52 autosomal SNPs. *International Congress Series*, 4, vol. 1288, no. 0, pp. 67-69 ISSN 0531-5131. DOI 10.1016/j.ics.2005.10.039.

SANCHEZ, J.J., BØRSTING, C., HALLENBERG, C., BUCHARD, A., HERNANDEZ, A. and MORLING, N., 2003. Multiplex PCR and minisequencing of SNPs—a model with 35 Y chromosome SNPs. *Forensic Science International*, 10/14, vol. 137, no. 1, pp. 74-84 ISSN 0379-0738. DOI 10.1016/S0379-0738(03)00299-8.

SANTOS, C., PHILLIPS, C., FONDEVILA, M., PORRAS-HURTADO, L., CARRACEDO, A., SOUTO, L. and LAREU, M.V., 2011. A study of East Timor variability using the SNPforID 52-plex SNP panel. *Forensic Science International-Genetics*, JAN 2011, vol. 5, no. 1, pp. E25-E26 ISSN 1872-4973. DOI 10.1016/j.fsigen.2010.03.006.

SCHWARK, T., MEYER, P., HARDER, M., MODROW, J. and VON WURMB-SCHWARK, N., 2012a. The SNPforID Assay as a Supplementary Method in Kinship and Trace Analysis. *Transfusion Medicine and Hemotherapy*, JUN 2012, vol. 39, no. 3, pp. 187-193 ISSN 1660-3796. DOI 10.1159/000338855.

SCHWARK, T., MEYER, P., HARDER, M., MODROW, J. and VON WURMB-SCHWARK, N., 2012b. The SNPforID Assay as a Supplementary Method in Kinship and Trace Analysis. *Transfusion Medicine and Hemotherapy*, JUN 2012, vol. 39, no. 3, pp. 187-193 ISSN 1660-3796. DOI 10.1159/000338855.

SEAH, L.H., JEEVAN, N.H., OTHMAN, M.I., JAYA, P., OOI, Y.S., WONG, P.C. and LEE, S.S., 2003. STR data for the AmpFlSTR identifiler loci in three ethnic groups (Malay, Chinese, Indian) of the Malaysian population. *Forensic Science International*, DEC 17 2003, vol. 138, no. 1-3, pp. 134-+ ISSN 0379-0738. DOI 10.1016/j.forsciint.2003.09.005.

SERRE, D. and PAABO, S.P., 2004. Evidence for gradients of human genetic diversity within and among continents. *Genome Research*, SEP 2004, vol. 14, no. 9, pp. 1679-1685 ISSN 1088-9051. DOI 10.1101/gn2529604.

SHARAFI FARZAD, M., TOMAS, C., BØRSTING, C., ZEINALI, Z., MALEKDOOST, M., ZEINALI, S. and MORLING, N., 2013. Analysis of 49 autosomal SNPs in three ethnic groups from Iran: Persians, Lurs and Kurds. *Forensic Science International: Genetics*, 7, vol. 7, no. 4, pp. 471-473 ISSN 1872-4973. DOI http://dx.doi.org/10.1016/j.fsigen.2013.04.001.

Snipper SNP-Indel application portal. Available from: http://mathgene.usc.es/snipper/.

SNPforID Consortium. *The SNPforID Browser.*

SOBRINO, B., BRIÓN, M. and CARRACEDO, A., 2005a. SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Science International*, 11/25, vol. 154, no. 2–3, pp. 181-194 ISSN 0379-0738. DOI 10.1016/j.forsciint.2004.10.020.

SOBRINO, B., BRIÓN, M. and CARRACEDO, A., 2005b. SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Science International*, 11/25, vol. 154, no. 2–3, pp. 181-194 ISSN 0379-0738. DOI 10.1016/j.forsciint.2004.10.020.

SUADI, Z., SIEW, L.C., TIE, R., HUI, W.B., ASAM, A., THIEW, S.H. and BOON, L.K., 2007. STR Data for the AmpFlSTR® Identifiler® Loci from the Three Main Ethnic Indigenous Population Groups (Iban, Bidayuh, and Melanau) in Sarawak, Malaysia. *Journal of Forensic Sciences*, vol. 52, no. 1, pp. 231-234 ISSN 1556-4029. DOI 10.1111/j.1556-4029.2006.00344.x.

The International HapMap Consortium, 2003. The International HapMap Project, vol. 426, no. 6968, pp. 789-796.

THORISSON, G.A., SMITH, A.V., KRISHNAN, L. and STEIN, L.D., 2005. The international HapMap project web site. *Genome Research*, vol. 15, no. 11, pp. 1592-1593.

THORISSON, G.A. and STEIN, L.D., 2003. The SNP Consortium website: past, present and future. *Nucleic Acids Research*, January 01, vol. 31, no. 1, pp. 124-127 DOI 10.1093/nar/gkg052.

TOMAS, C., AXLER-DIPERTE, G., BUDIMLIJA, Z.M., BØRSTING, C., COBLE, M.D., DECKER, A.E., EISENBERG, A., FANG, R., FONDEVILA, M., FRISK FREDSLUND, S., GONZALEZ, S., HANSEN, A.J., HOFF-OLSEN, P., HAAS, C., KOHLER, P., KRIEGEL, A.K., LINDBLOM, B., MANOHAR, F., MAROÑAS, O., MOGENSEN, H.S., NEUREUTHER, K., NILSSON, H., SCHEIBLE, M.K., SCHNEIDER, P.M., SONNTAG, M.L., STANGEGAARD, M., SYNDERCOMBE-COURT, D., THACKER, C.R., VALLONE, P.M., WESTEN, A.A. and MORLING, N., 2011. Autosomal SNP typing of forensic samples with the GenPlex™ HID System: Results of a collaborative study. *Forensic Science International: Genetics*, 11, vol. 5, no. 5, pp. 369-375 ISSN 1872-4973. DOI 10.1016/j.fsigen.2010.06.007.

TOMAS, C., DIEZ, I.E., MONCADA, E., BORSTING, C. and MORLING, N., 2013. Analysis of 49 autosomal SNPs in an Iraqi population. *Forensic Science International-Genetics*, JAN 2013, vol. 7, no. 1, pp. 198-199 ISSN 1872-4973. DOI 10.1016/j.fsigen.2012.05.004.

TOMAS, C., SANCHEZ, J.J., AURELIO CASTRO, J., BORSTING, C. and MORLING, N., 2010. Forensic usefulness of a 25 X-chromosome single-nucleotide polymorphism marker set. *Transfusion*, OCT 2010, vol. 50, no. 10, pp. 2258-2265 ISSN 0041-1132. DOI 10.1111/j.1537-2995.2010.02696.x.

TOMAS, C., STANGEGAARD, M., BØRSTING, C., HANSEN, A.J. and MORLING, N., 2008. Typing of 48 autosomal SNPs and amelogenin with GenPlex SNP genotyping system in forensic genetics. *Forensic Science International: Genetics*, 12, vol. 3, no. 1, pp. 1-6 ISSN 1872-4973. DOI 10.1016/j.fsigen.2008.06.007.

VAINER, M., ENAD, S., DOLNIK, V., XU, D., BASHKIN, J., MARSH, M., TU, O., HARRIS, D.W., BARKER, D.L. and MANSFIELD, E.S., 1997. Short Tandem Repeat Typing by Capillary Array Electrophoresis: Comparison of Sizing Accuracy and Precision Using Different Buffer Systems. *Genomics*, 4/1, vol. 41, no. 1, pp. 1-9 ISSN 0888-7543. DOI 10.1006/geno.1997.4608.

VULLO, C., BOROSKY, A., CATELLI, M., ROMANINI, C., FONDEVILA, M., SANTOS, C., PEREIRA, R. and GUSMAO, L., 2011. Population data for 38 autosomal insertion/deletion (InDels) and 50 SNPS polymorphisms in Argentinean population. *Forensic Science International: Genetics Supplement Series*, 12, vol. 3, no. 1, pp. e419-e420 ISSN 1875-1768. DOI http://dx.doi.org/10.1016/j.fsigss.2011.09.071.

WALSH, S., LINDENBERGH, A., ZUNIGA, S.B., SIJEN, T., DE KNIJFF, P., KAYSER, M. and BALLANTYNE, K.N., 2011. Developmental validation of the IrisPlex system: Determination of blue and brown iris colour for forensic intelligence. *Forensic Science International-Genetics*, NOV 2011, vol. 5, no. 5, pp. 464-471 ISSN 1872-4973. DOI 10.1016/j.fsigen.2010.09.008.

WALSH, S., LIU, F., WOLLSTEIN, A., KOVATSI, L., RALF, A., KOSINIAK-KAMYSZ, A., BRANICKI, W. and KAYSER, M., 2013. The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Science International-Genetics*, JAN 2013, vol. 7, no. 1, pp. 98-115 ISSN 1872-4973. DOI 10.1016/j.fsigen.2012.07.005.

WENZ, H.M., 1994. Capillary Electrophoresis as a Technique to Analyze Sequence-Induced Anomalously Migrating Dna Fragments. *Nucleic Acids Research*, SEP 25 1994, vol. 22, no. 19, pp. 4002-4008 ISSN 0305-1048. DOI 10.1093/nar/22.19.4002.

WENZ, H.M., ROBERTSON, J.M., MENCHEN, S., OAKS, F., DEMOREST, D.M., SCHEIBLER, D., ROSENBLUM, B.B., WIKE, C., GILBERT, D.A. and EFCAVITCH, J.W., 1998. High-precision genotyping by denaturing capillary electrophoresis. *Genome Research*, JAN 1998, vol. 8, no. 1, pp. 69-80 ISSN 1054-9803.

WESTRING, C.G., KRISTINSSON, R., GILBERT, D.M. and DANIELSON, P.B., 2007. Validation of reduced-scale reactions for the Quantifiler (TM) human DNA kit. *Journal of Forensic Sciences*, SEP 2007, vol. 52, no. 5, pp. 1035-1043 ISSN 0022-1198. DOI 10.1111/j.1556-4029.2007.00525.x.

WETTON, J.H., TSANG, K.W. and KHAN, H., 2005. Inferring the population of origin of DNA evidence within the UK by allele-specific hybridization of Y-SNPs. *Forensic Science International*, 8/11, vol. 152, no. 1, pp. 45-53 ISSN 0379-0738. DOI http://dx.doi.org/10.1016/j.forsciint.2005.03.009.

YONG, R.Y.Y., LEE, L.K.H. and YAP, E.P.H., 2006. Y-chromosome STR haplotype diversity in three ethnic populations in Singapore. *Forensic Science International*, 6/2, vol. 159, no. 2–3, pp. 244-257 ISSN 0379-0738. DOI http://dx.doi.org/10.1016/j.forsciint.2005.05.010.

ZAINUDDIN, Z. and GOODWIN, W., 2004. Mitochondrial DNA profiling of modern Malay and Orang Asli populations in peninsular Malaysia. *International Congress Series*, 4, vol. 1261, no. 0, pp. 428-430 ISSN 0531-5131. DOI http://dx.doi.org/10.1016/S0531-5131(03)01725-4.

Zarinah Mahari., 2011. *Demographic Transition in Malaysia: The Changing Roles of Women.* Wan Roslida Othman, Noor Masayu Mhd Khalili, Sabariah Esa and Nadia Miskiman ed., 15th Conference of Commonwealth Statisticians, New Delhi, India, 7-10 February 2011.

ZUBAKOV, D., BOERSMA, A.W.M., CHOI, Y., VAN KUIJK, P.F., WIEMER, E.A.C. and KAYSER, M., 2010. MicroRNA markers for forensic body fluid identification obtained from microarray

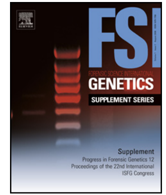
screening and quantitative RT-PCR confirmation*. International Journal of Legal Medicine*, MAY 2010, vol. 124, no. 3, pp. 217-226 ISSN 0937-9827. DOI 10.1007/s00414-009-0402-3.

Forensic Science International: Genetics Supplement Series 4 (2013) e178–e179

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

journal homepage: www.elsevier.com/locate/FSIGSS



SNP genotyping of forensic casework samples using the 52 SNPforID markers

S. Alimat [a,b], S. Hadi [a], W. Goodwin [a,*]

[a] School of Forensic and Investigative Sciences, University of Central Lancashire, Preston, United Kingdom
[b] Forensic Division, Department of Chemistry Malaysia (KIMIA), Ministry of Science and Technology and Innovation, Malaysia

ARTICLE INFO

Article history:
Received 4 September 2013
Accepted 2 October 2013

Keywords:
DNA profiling
Degraded DNA
Single nucleotide polymorphisms
SNPforID

ABSTRACT

The analysis of degraded DNA is one of the biggest challenges in forensic casework. SNPs, which can be amplified using small amplicons, have previously been successfully applied to the profiling of forensic evidence that could not be analyzed using conventional STRs. Here we selected the 52 SNPforID SNP markers, with amplicons that ranged in size from 59 bp to 115 bp, and used them to profile a range of casework samples from Malaysia. DNA degradation is a common problem in Malaysia due to the high temperatures and humidity. To carry out the study we modified the 52 SNPforID markers into four 13-plex SNaPshot assays to enable easier interpretation of profiles on the ABI PRISM® 310 and 3500.

Fifty-one crime samples comprising bloodstains on cloth, swabs, and a mat and 2 swabs of trace DNA from 10 crime scenes in Malaysia were profiled after DNA extraction using a phenol–chloroform method. The samples were also subjected to STR analysis using the Powerplex® 16 system (Promega), which resulted in only 17 full profiles and 9 partial profiles; using SNPs, 36 full profiles and 5 partial profiles could be generated.

© 2013 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

SNPs have received a lot of attention in recent years due to their abundance throughout the human genome and have been widely used in medical studies. Their use in forensic DNA analysis is limited, but is becoming more widespread. Applying SNPs to the analysis of degraded DNA is of particular interest as SNPs have the advantage over short tandem repeats (STRs) that the amplicon size is generally much shorter and therefore they have the potential to be successfully profiled in degraded samples. Crime scene samples in Malaysia often display signs of degradation as the environmental is hot and humid.

The main objective of the study was to assess the success rate in profiling crime scene samples using the SNPs designed by the SNPforID project [1,2] in comparison to a conventional STR system.

2. Process

Fifty-one real crime samples (2 swabs and bloodstains from clothing and a mat) were provided by Forensic DNA/Serology Section, Department of Chemistry, Malaysia. DNA was extracted using a phenol–chloroform method and then was quantified using Quantifiler® Kit. All crime scene samples had been processed and reported and the cases processed through the courts in Malaysia prior to this work.

In this study, four sets of 13-PCR & SBE (single base extension)-plexes were designed and evaluated before analyzing the crime samples. All samples (1 μl of DNA input ranged from undetermined up to 100 ng) were genotyped using four separate 13-plex PCR assays. PCR products were purified with ExoSAP enzymes. The 13-plex SBE amplification of the purified samples was carried out using SNaPshot Multiplex® kit. The SBE products were purified with SAP enzyme. STR analysis using the Powerplex® 16 System

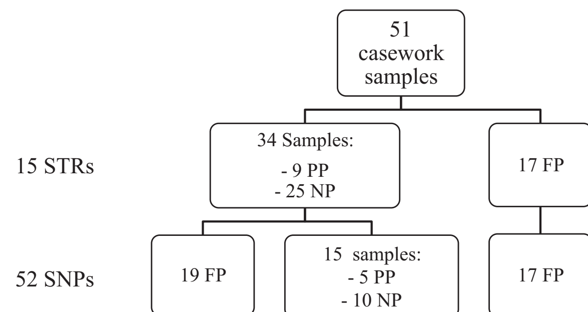


Fig. 1. Success rate of STR and SNP profiles developed from casework samples. FP = full profile; PP = partial profile; NP = no profile. A SNP partial profile required more than 26 markers to have been called and for STR partial profiles seven or more loci were needed.

* Corresponding author. Tel.: +44 1772 894254; fax: +44 1772 894981.
E-mail addresses: SAlimat@uclan.ac.uk (S. Alimat), SHadi@uclan.ac.uk (S. Hadi), whgoodwin@uclan.ac.uk (W. Goodwin).

1875-1768/$ – see front matter © 2013 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.fsigss.2013.10.092

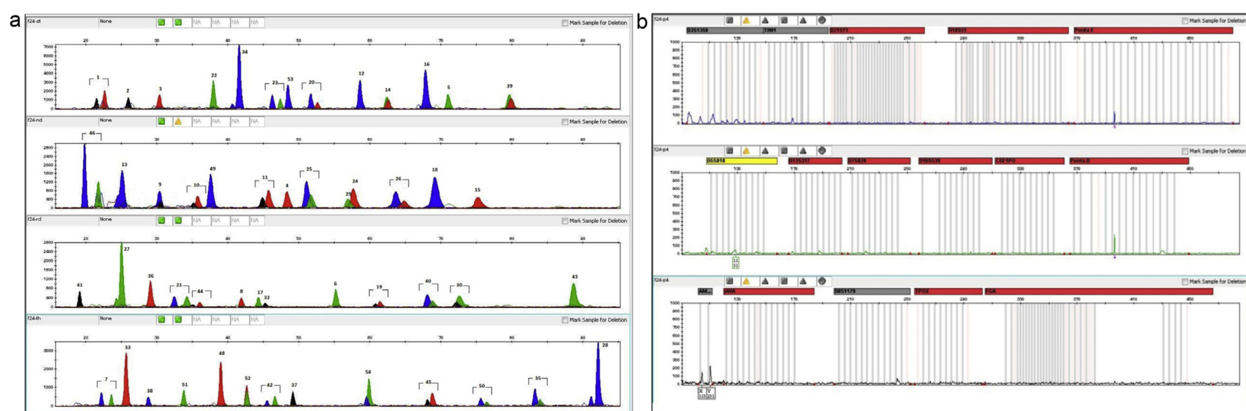


Fig. 2. Profiling of a crime scene sample using (a) STR markers and (b) SNP markers. A full profile could be generated using the SNP markers whereas no loci were successful amplified using STRs.

(Promega) was also carried out on all the samples. Samples from both analyses were subjected to capillary electrophoresis using the Applied Biosystems 310 and 3500 Genetic Analyzers and analyze using GeneMapper software with peak thresholds set to a minimum of 120 RFUs (blue colour), 60 RFUs (green colour) and 30 RFUs (yellow, red and orange colours). The criteria for acceptable peak height ratio for heterozygotes allele calls and homozygotes allele calls for each locus were as previously described [1].

3. Results

Out of 51 selected samples; 17 full STR profiles were generated using conventional STR analysis; of the remaining 34 samples 9 gave partial profiles with 7 or more loci successfully analyzed; 25 could not be amplified at all. In comparison, with 52 SNP markers, 36 full profiles were successfully amplified from the samples and only 15 were partially or not amplified at all (with 5 partial profiles where more than 26 loci had been amplified) and no DNA profile was obtained for 10 samples (Fig. 1)

An example of profiles from a poor quality sample is given in Fig. 2.

4. Discussion

We have demonstrated with casework samples that in some instances the SNPs can generate full profiles from DNA extracts that yielded no or few STR loci. The increased success rate will be partially due to the reduced amplicon sizes and may also be due to greater resistance to inhibitors as the chemistries used to amplify the two types of marker were slightly different. In the example shown in Fig. 2 we would expect some of the STR loci which have shorter amplicons to be profiled and inhibition may have impacted more on the STR typing.

The increase in success rate is notable; from approximately 50% of samples profiled using STRs to over 80% successfully profiled using SNPs. Further work needs to be done to assess whether the increased success rate can be achieved using less labour intensive systems, such as mini STRs [3] and INDELS [4].

Role of funding

This study was supported by the Ministry of Science, Technology and Innovation (MOSTI) in Malaysia.

Conflict of interest

None.

Acknowledgements

We thank the Director General, Section Head and the staff at the DNA Laboratory of the Forensic Division, Department of Chemistry and the Ministry of Science, Technology and Innovation (MOSTI) in Malaysia. Special appreciations to the UCLan Forensic Genetics group for their valuable advice and assistance in the laboratory.

References

[1] J.J. Sanchez, et al., A multiplex assay with 52 single nucleotide polymorphisms for human identification, Electrophoresis 27 (2006) (2006) 1713–1724.

[2] The SNPforID Browser (spsmart.cesga.es/snpforid.php).

[3] M. Alenizi, W. Goodwin, S. Hadi, H. Alenizi, K. Altamar, M. Alsikel, Concordance between the AmpFℓSTR® MiniFiler™ and AmpFℓSTR® Identifiler® PCR Amplification Kits in the Kuwaiti population, J. Forensic Sci. 54 (2009) 350–352.

[4] C. Romanini, M.L. Catelli, A. Borosky, R. Pereira, M. Romero, M. Salado Puerto, C. Phillips, M. Fondevila, A. Freire, C. Santos, A. Carracedo, M.V. Lareu, L. Gusmao, C.M. Vullo, Typing short amplicon binary polymorphisms: Supplementary SNP and Indel genetic information in the analysis of highly degraded skeletal remains, Forensic Sci. Int.: Genet. 6 (4) (2012) 469–476.

Muhammad S. Nazir[1,2]
Sasitaran Iyavoo[1]
Sharizah Alimat[1]
Nathalie Zahra[1]
Sheikha H. Sanqoor[1,3]
Judith A. Smith[1]
Colin Moffatt[1]
Will Goodwin[1]

[1]School of Forensic and Investigative Sciences, University of Central Lancashire, Preston, UK
[2]University of Modern Sciences, Dubai, UAE
[3]General Department of Forensic Science and Criminology, Biology and DNA Section, Dubai Police General H.Q., Dubai, UAE

Received May 22, 2013
Revised September 15, 2013
Accepted September 16, 2013

Research Article

Development of a multiplex system to assess DNA persistence in taphonomic studies

In this study, we have developed a PCR multiplex that can be used to assess DNA degradation and at the same time monitor for inhibition: primers have been designed to amplify human, pig, and rabbit DNA, allowing pig and rabbit to be used as experimental models for taphonomic research, but also enabling studies on human DNA persistence in forensic evidence. Internal amplified controls have been added to monitor for inhibition, allowing the effects of degradation and inhibition to be differentiated. Sequence data for single-copy nuclear recombination activation gene (RAG-1) from human, pig, and rabbit were aligned to identify conserved regions and primers were designed that targeted amplicons of 70, 194, 305, and 384 bp. Robust amplification in all three species was possible using as little as 0.3 ng of template DNA. These have been combined with primers that will amplify a bacterial DNA template within the PCR. The multiplex has been evaluated in a series of experiments to gain more knowledge of DNA persistence in soft tissues, which can be important when assessing what material to collect following events such as mass disasters or conflict, when muscle or bone material can be used to aid with the identification of human remains. The experiments used pigs as a model species. When whole pig bodies were exposed to the environment in Northwest England, DNA in muscle tissue persisted for over 24 days in the summer and over 77 days in the winter, with full profiles generated from these samples. In addition to time, accumulated degree days (ADD) were also used as a measure that combines both time and temperature—24 days was in summer equivalent to 295 ADD whereas 77 days in winter was equivalent to 494 ADD.

Keywords:
DNA degradation / DNA persistence / DNA profiling / Forensic genetics
DOI 10.1002/elps.201300240

Additional supporting information may be found in the online version of this article at the publisher's web-site

1 Introduction

Different experimental approaches have been utilized to estimate the level of DNA degradation in postmortem tissue and body fluids. These include comet assays [1], the development of species-specific PCR primers and agarose gel electrophoresis [2–5], species-specific primers and real-time PCR [6–8], Southern blotting and hybridization [9, 10], competitive PCR [11], and profiling of DNA extracts with polymorphic markers systems [12–14].

Understanding DNA degradation is important when undertaking forensic analysis [15]. For example, the level of degradation dictates which tissues to sample from human remains when using DNA to assist with the identification [16, 17]. Bone and teeth are preferred over soft tissues when dealing with highly degraded remains; however, processing of hard tissues takes longer and is more expensive than when using muscle tissues. Assessing the level of degradation can also be important in deciding which DNA profiling systems are most appropriate for the analysis of different sample types. While most laboratories will have a standard STR-based profiling system that will typically amplify DNA up to 400–500 bp, highly degraded samples may benefit from analysis using specialized assays that either amplify STR markers using short amplicons [18–20], target SNPs [21–24], or insertion-deletion polymorphisms (indels) [25, 26].

Correspondence: Dr. Will Goodwin, School of Forensic and Investigative Sciences, University of Central Lancashire, PR1 2HE Preston, UK
E-mail: whgoodwin@uclan.ac.uk

Abbreviations: ADD, accumulated degree days; **IAC**, internal amplification control; **RAG-1**, recombination activation gene

Colour Online: See the article online to view Figs. 1 and 2 in colour.

© 2013 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

In previous studies, the reported persistence of DNA in soft muscle has varied: experiments in Western Australia using pig muscle from whole animals could amplify DNA after ten days in winter [3], in Thailand pig muscle yielded DNA for up to four days when exposed to the environment (eight days when immersed in water) [5], rat brain tissue yielded DNA when stored at 20°C for up to 42 days [8], human brain and thyroid tissue yielded DNA for over 100 days when stored in a container or immersed in water at 21°C [14]. Of the above studies, only Larkin et al. [3] took into account the level of DNA degradation in relation to both time and temperature (accumulated degree days (ADD)) for the purpose of estimating postmortem interval [3, 27–29].

Here we describe the development of a PCR multiplex that targets loci with amplicons over a similar range to that seen in many commercial PCR amplification kits used for forensic DNA profiling. We have combined the multiplex with previously described internal amplification controls (IACs) that can monitor for PCR inhibition [30, 31]. We have used the optimized multiplex to work with human, pig, and rabbit DNA, allowing it to be used as a tool to assess the persistence/degradation of DNA in soft muscle tissue using pigs or rabbits as an experimental model. We describe an application of the multiplex assessing the persistence of DNA in pig muscle tissue in a series of field experiments.

2 Materials and methods

2.1 Design of PCR primers

Sequence data for a nuclear recombination activation gene (RAG-1) from human, rabbit, and pig were downloaded from GenBank and aligned using Bioedit (http://www.mbio.ncsu.edu/bioedit/bioedit.html) to identify conserved regions: primers that would amplify 70, 194, 305, and 384 bp amplicons from the three species were identified (Table 1). These primers were designed using the publicly available software Primer3 [32, 33] and the oligonucleotide property calculator software [34].

Primer pairs with 5′ fluorescein-labeled or ROX-labeled forward primers were synthesized (Life Technologies™, UK) and were delivered desalted and lyophilized: 100 μM stock solutions were prepared by adding the appropriate volume of 1× TE buffer (0.1 M Tris HCl, 0.01 M EDTA, pH 8.0; Sigma, UK) and stored at −20°C, while an aliquot of a 10 μM working solution was kept at 4°C.

2.2 DNA samples

DNA extraction was carried out using DNeasy® Blood and Tissue kit (Qiagen, UK) according to the manufacturer's instructions using 25–30 mg of tissue. Pig and rabbit positive control DNA samples were extracted from soft muscle tissue samples collected at day 0.

2.3 DNA quantification

Extracts used in the sensitivity study were quantified in triplicate using real-time PCR of the 70 bp fragment with two unlabeled primers. Go*Taq*® qPCR Master Mix (Promega®, UK) was used following the manufacturer's recommended protocol with an ABI 7500 real-time PCR machine (Life Technologies™).

Field samples were extracted as above, but quantified using the Quant-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies™). Samples were prepared by mixing 3 μL DNA with 147 μL 1× TE and 150 μL PicoGreen and pipetted (300 μL) into microplate wells—all samples were prepared in triplicate. Fluorescence was determined using TECAN GEnios Pro plate reader at absorption and emission wavelengths of 485 and 535 nm, respectively. Data were obtained in the form of relative fluorescent unit.

2.4 Multiplex PCR

The primer mix and thermal cycler conditions were prepared according to the optimized PCR condition (Table 1). In a total reaction volume of 10.0 μL, 5.0 μL 2× Platinum Multiplex PCR Master Mix (Applied Biosystems), 0.6 μL of primers mix, 2.5 μL of dH_2O, 1 μL of DNA template, and 1 μL of IAC template, prepared as described by Zahra et al. [30], were added. PCR was carried out in a thermal cycler GeneAmp® 2700 (Life Technologies™) using following conditions: 95°C for 5 min; 30 cycles of 94°C for 1 min, 60°C for 1 min, 72°C for 1 min, and a final extension at 72°C for 10 min.

For sensitivity studies dilutions (2.80, 0.93, 0.31, 0.10, and 0.03 ng) were made for human, pig, and rabbit day DNA. Human genomic DNA G3041 (273 ng/μL) was obtained from Promega®.

2.5 Capillary electrophoresis

Each sample for fragment analysis was prepared by adding 1 μL of PCR product to 10 μL of Hi-Di™ formamide (Life Technologies™) containing 0.7 μL GS500 ROX size standard (Life Technologies™). The samples were heated at 95°C for 5 min and snap-cooled for at least 3 min.

Optimization of the multiplex PCR was performed using an ABI 310 Prism® Genetic Analyzer (Life Technologies™). DNA fragment analysis was carried out using a 47 cm-long capillary using POP-4™ polymer (Life Technologies™). Electrophoresis running buffer was used at 1× concentration. The GSPOP 4 (1 mL) F.md4 run module with dye set DS-32 (filter set F) was used with the following parameters: run temperature 60°C, syringe pump time 150 s, prerun voltage 15 kV, prerun time 120 s, injection time 5 s, injection voltage 15 kV, run voltage 15 kV, and run time 30 min.

Due to the ability to increase throughput DNA fragment, analysis for the field samples was performed using an ABI 3500 Genetic Analyzer (Life Technologies) using a 50 cm

© 2013 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 1. Shown below are four PCR primer sets for conserved regions of pig, rabbit, and human

Nuclear gene	Forward and reverse primers (5′-3′)	Primer conc. (μM)	Amplicon (bp)
RAG-1 NCBI reference M77666.1	CCT CAA AGT CAT GGG CAG C	0.08	70
	GAC TCT CCA GGT CAG TAG G	(0.05)	
	GCT GTT TGC TTG GCC ATC CG	0.19	194
	GTG CTG GAA GAC ACA TTC TTC	(0.1)	
	ATG AGG TCT GGC GTT CCA AC	0.15	305
	TGG TCA TGA GCT TCC TGG CA	(0.15)	
	GAG CAA TCT CCA GCA GTC CT	0.56	384
	GCT AAA CTT CCT GTG CAT GA	(0.4)	
IAC 90	CTG TCA AAT CTA AAC ACC CTG ATG CG[a)]	0.60[a)]	90
	GTC AGC TTG CAT AAT ATC GAG ATA ACG C	0.15	
IAC 410	CTG TCA AAT CTA AAC ACC CTG ATG CG[a)]	0.60[a)]	410
	GTA CAA TGT TGA CGT TCC TCG CTG	0.45	

The sequence and concentration of the IAC primers are also given. The primer concentrations shown in brackets are used when the IAC primers are not included.
a) The same forward primer is used for both the IAC 90 and 410; 0.6 μM is the total primer concentration.

capillary array and 3500 POP-6™ polymer. The Fragment-Analysis50_POP6 run module was used in combination with the dye set DS-32 (filter set F with the following parameters: run temperature 60°C, prerun voltage 15 kV, prerun time 180 s, injection time 10 s, injection voltage 1.6 kV, run voltage 15 kV, and run time 2700 s).

The data obtained from CE were analyzed using GeneMapper™ ID version 3.2 (ABI 310) and GeneMapper® Software v4.1 (ABI 3500) (both Life Technologies™).

2.6 Simulated DNA degradation

The exposure of forensic evidence samples to environmental insult, fragmentation of full-length genomic DNA, and the reduction of overall concentration of amplifiable DNA are observed. Therefore, the effect of DNA degradation on the amplification efficiency of the multiplex (4-plex) was examined according to the protocol described by Swango et al. [35]. A degradation series was prepared by digesting high molecular weight (177 μg/mL) DNA with DNase I (Applied Biosystems) for a progressive length of time. A reaction containing 15.7 μg DNA, 10× DNase buffer (100 mM Tris, 25 mM $MgCl_2$, 5 mM $CaCl_2$, pH 7.6), and nuclease-free water to bring the reaction volume to 110 μL, was prepared. Ten microliters were removed and used as a control sample. DNA in the remaining solution was digested by adding 1.25 μL of DNase I (2 U/μL), and then the mixture was incubated at 24°C. From this solution, 10 μL was removed at specific time points (5, 10, 15, 20, 30, 45, 60, 90, 120, and 180 min), DNase I activity was stopped by mixing 2 μL of DNase inactivation reagent, and incubating at 65°C for 15 min. After incubation, samples were centrifuged at 10 000 × *g* for 2 min and supernatant was transferred to a fresh tube. The degree of DNA degradation was assessed by running 2 μL of digested products on 2% agarose gel using 100 bp DNA ladder (BioLabs®, UK). DNA quantification was performed using Human Quantifiler® kit (Life Technologies™) according to the manufacturer's instruction.

2.7 Field studies

In order to assess the degradation of DNA in soft muscle tissue whole pig carcasses were laid out at the University of Central Lancashire's Taphonomic Research in Anthropology: Centre for Experimental Studies (TRACES), which comprises 13 acres of field facilities in the Northwest of England [36]; carcasses were covered by a wire mesh cage to prevent scavenging by large mammals. Field experiments were carried out in February–May 2010, May–June 2010, and September–October 2010. Muscle samples of approximately 1 g were taken from the upper hind limb at regular intervals until either no DNA could be recovered (this was based on an assessment of the DNA extract on an agarose gel and was only relevant for the February series) or no muscle tissue remained for analysis (as occurred with both the May and September experiments). The muscle tissues remain moist throughout all experiments described here and samples were taken from below the tissue surface, avoiding the complications that could be introduced by muscle tissues becoming desiccated, which would have enhanced the preservation of DNA. In a preliminary series of experiments related to this study, different regions of muscles were sampled: the hind and fore legs, both upper and lower (i.e. in contact with the ground) and the neck, no significant difference was seen in the rates of degradation [37]. Samples were collected in all cases from three separate pig carcasses and were transported on ice and frozen at −20°C until DNA was extracted. Ambient temperatures were measured at the field site using HDT-200 and HDT-250 USB data loggers (Thermosense, UK). When analyzing the field samples with the multiplex, 0.6 ng of DNA template were added to each PCR and the CE undertaken using a 3500 Genetic Analyzer.

© 2013 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

3 Results

3.1 Selection of loci and primer design

Single-copy RAG-1 is 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. The RAG-1 gene is found throughout higher vertebrates and comprises a 3.1 kb exon without introns. It evolves slowly, has minimal saturation at the third position of codons, and a low frequency of indels [38–40]. Multiple conserved regions could be identified in the alignment between human, pig, and rabbit sequence, allowing a range of amplicons sizes from the same gene (Supporting Information Fig. 1). Additional sites were available allowing amplicons be added at a later stage, if desired.

3.2 Multiplex sensitivity and specificity

The four primer pairs were found to be optimum at working concentrations of 0.05, 0.1, 0.15, and 0.4 μM for 70, 194, 305, and 384 bp, respectively (Table 1). The $MgCl_2$ concentration was kept at 2.5 mM for all PCR. The optimized 4-plex PCR was assessed for any nonspecific amplification that would lead to extra peaks and could interfere with target peaks. Positive control DNA samples of all three species were evaluated on ABI Prism® 310 and 3500 Genetic Analyzers.

DNA fragment analysis for multiplex PCR was carried out using serial dilutions of positive control DNA samples of human (Promega®), rabbit, and pig extracted from day 0 postmortem soft muscle tissues and quantified using real-time PCR. The 4-plex PCR was found to work efficiently in triplicate samples of all three species between 2.80 and 0.1 ng of DNA template (Supporting Information Fig. 2); no extra peaks were observed in any of the amplifications and full amplification was observed in all three species with template amounts as low as 0.3 ng; when the template amount dropped below 0.3 ng allele dropout, which is typical of low-template PCR [41, 42], was observed in some profiles. The multiplex was also tested with DNA from insects that would commonly be found on the carcasses [43]; no amplification was seen with DNA from *Calliphora vomitoria*, *Protophormia terraenovae*, *Lucilia sericata*, *Lucilia caesar*, or *Lucilia illustris*. Addition of the IAC template and primers did not alter the sensitivity of specificity of the multiplex.

To assess the ability of the multiplex (4-plex) assay to quantify DNA in degraded samples, a ten-point degradation series was prepared by treating aliquots of high molecular weight genomic DNA (Promega®) with DNase I for increasing periods of time ranging from 5 to 180 min. The result was, as expected, a degradation series exhibiting incremental increase in the extent of DNA degradation. Full 4-plex profiles were obtained reproducibly until 30 min of DNase I digestion, but with a decline in overall peak height and a relative decline in the amplification of large loci. There was complete failure of amplification in all DNA samples digested with DNase I for a period of 45 min or above (Supporting Information Fig. 3).

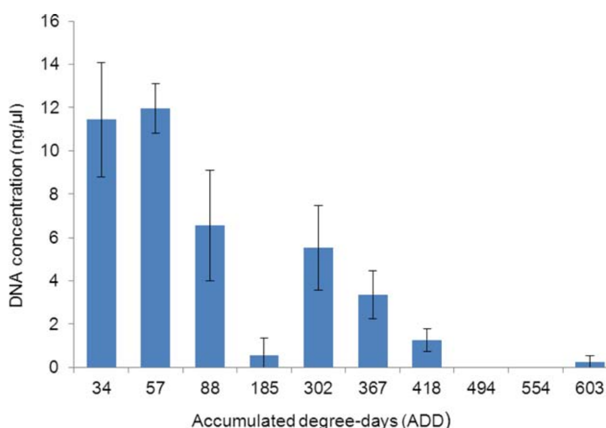


Figure 1. Shown above is the PicoGreen®-based DNA quantification of samples collected from pig carcasses in February 2010. The error bars indicate the standard error from three replicates. A total of 200 μL was extracted from each sample and 3 μL was used for quantification.

The standard deviation of the amplicon sizes for multiple runs of the same and different samples was below 1 bp for all amplicons. The sizing of fragments varied in accuracy: 70 bp (64.32 ± 0.69 bp); 194 bp (194.54 ± 0.04 bp), 305 bp (305 ± 0.5 bp), and 384 bp (383.9 ± 0.05 bp) on the ABI 310 Genetic Analyzer. Slightly different values were seen when using the 3500 Genetic Analyzer; 70 bp (65.17 ± 0.02 bp), 194 bp (195.66 ± 0.03 bp), 305 bp (307.38 ± 0.60 bp), and 384 bp (384.69 ± 0.15 bp).

3.3 Taphonomic study using field DNA samples

Three separate repeats of an experiment to assess the persistence of DNA in soft tissue were undertaken over the course of a year (2010), with experiments covering winter–spring, spring–summer, and autumn. Sampling time points varied depending on the season and the temperatures: in May samples were taken at between three and six day intervals whereas in February samples were only taken approximately once a week.

The breakdown of soft tissues was much more rapid, as would be expected, in summer and sampling was not possible after day 27 in the May–June experiment. Following collection of soft muscle tissue DNA was extracted and quantified using PicoGreen®. The values obtained from the quantification were used as a guide for adding DNA to the multiplex. An example of the quantification results from the February series is shown in Fig. 1. PicoGreen® is nonspecific and so will quantify bacterial and fungal DNA. However, this did not appear to confound the quantification of the endogenous DNA as the values went down to zero over time, before then increasing as bacterial and fungus colonized the remaining tissues. Based

© 2013 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

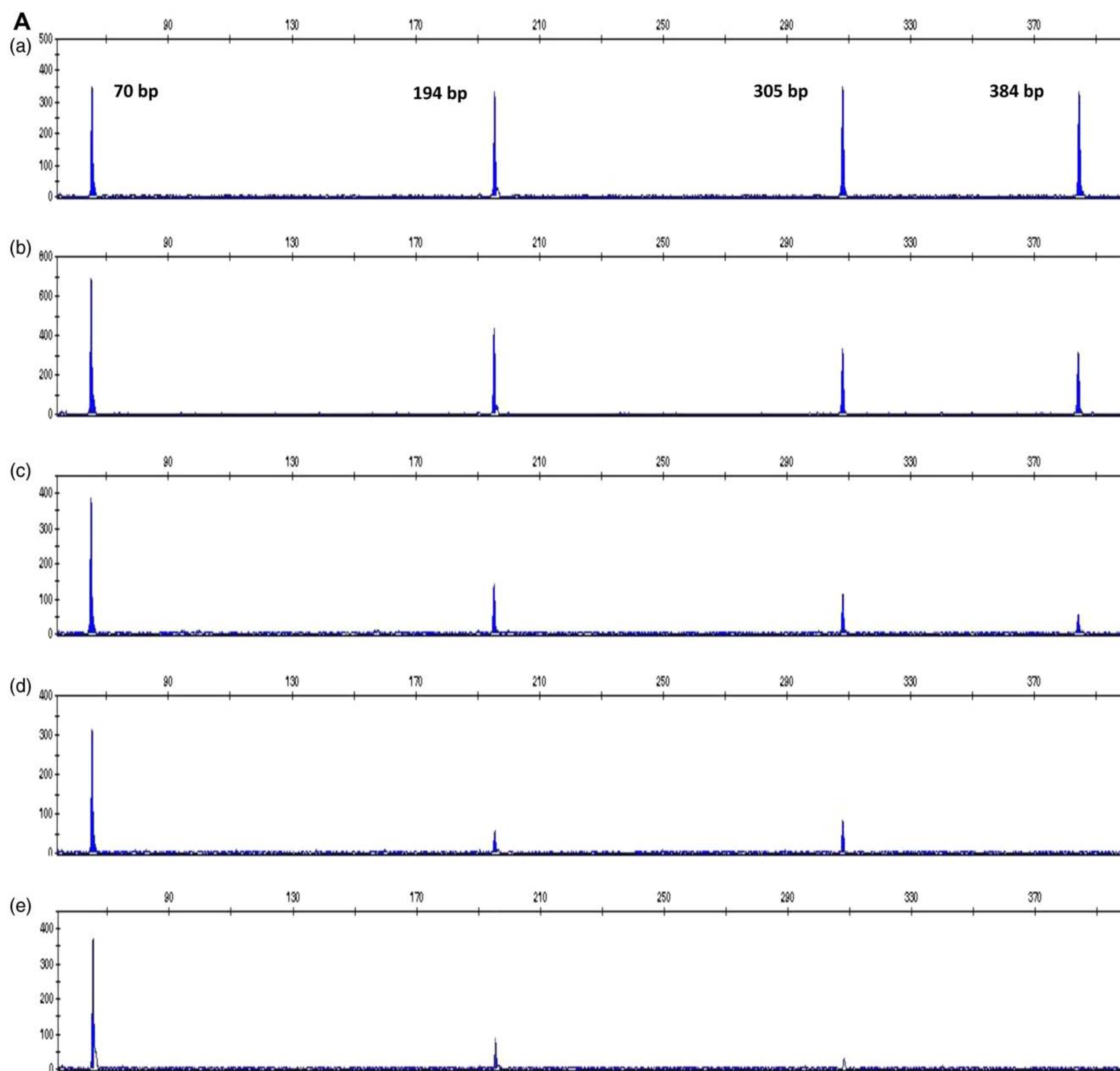


Figure 2. (A) Examples of electropherograms generated from 1 μL of DNA extract (from a total of 200 μL) extracted from pig soft muscle tissues collected during May 2010 project at (b) 159 ADD, (c) 203 ADD, (d) and (e) 338 ADD. Human positive control (a) is also shown (0.5 ng). The amount of DNA in 1 μL of each extract was quantified using real-time PCR: (b) 0.82 ng, (c) 0.73 ng, (d) 0.32 ng, (e) 0.59 ng. (B) Electropherograms with the same samples as in (A) with the addition of the IAC 90 and 410 internal amplification controls.

on the quantifications, 0.6 ng was added to the PCR multiplex and the absence or presence of each of the amplicons was noted. The PCRs were generally well balanced, which indicated that the quantification using PicoGreen® was reasonably accurate. An example of electropherograms taken from the May series is shown in Fig. 2 (February and September are shown in Supporting Information Figs. 4 and 5)—input DNA was kept constant for these using 1 μL of extract in each PCR (i.e. 0.5% of the total extract) and the amount of DNA added was assessed by real-time PCR. As time/ADD increase, the larger amplicons show a relative decline in abundance. This is expected, but somewhat surprisingly amplicons can still be amplified up to day 90/603 ADD in the February experiment. The IAC primers shown in Fig. 2B produce two peaks of 90 and 410 bp; these are well balanced in all of the electropherograms, indicating the absence of PCR inhibitors (had PCR inhibitors been present then we would expect the 410 bp IAC to decrease with respect to the 90 bp IAC). Table 2 contains the data from all three field experiments—0.6 ng was used as template in each of the reactions. As expected, DNA persists

© 2013 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

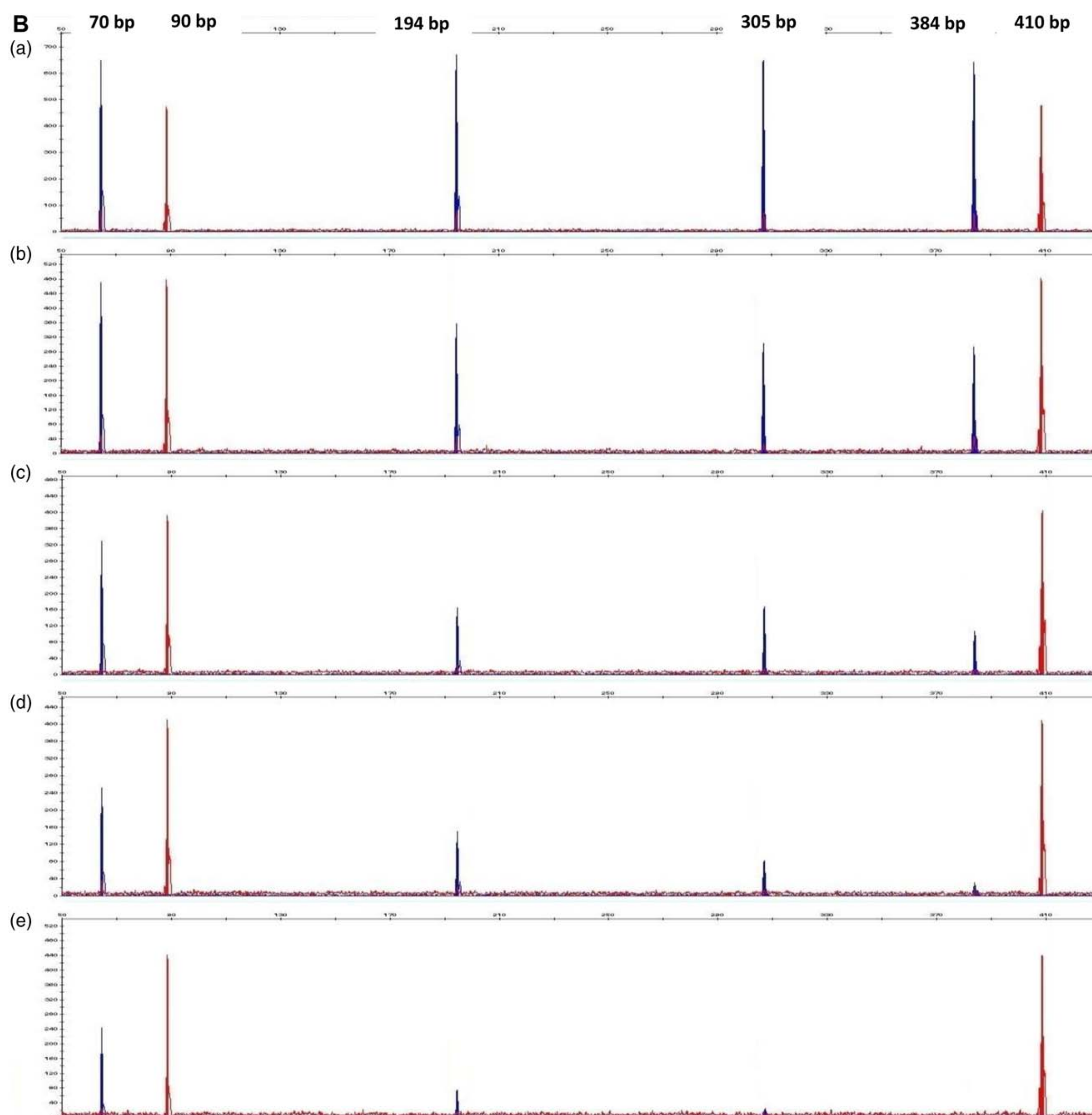


Figure 2. Continued.

for fewer days as the average temperature increases. However, full profiles could be obtained from all samples in the May series (average daily temperature 12.3°C) up to and including day 24 (ADD 295) and in September series (average daily temperature 11.4°C) up to and including day 30 (ADD 342). We could have potentially increased the number of loci amplified in some of the degraded samples by adding more template DNA, but in the context of this series of experiments were more interested in the general pattern of degradation and providing a conservative estimate of DNA persistence—with casework samples it may be appropriate to carry out further analysis.

The data were analyzed using the generalized linear mixed model fit by the Laplace approximation to measure analysis of deviance using the software R (R Development Core Team. http://www.R-project.org). The analysis demonstrated that the relationship between ADD and amplification success was not different for the 305 and 384 bp amplicons ($Z = 0.96$, $p = 0.76$), but was different for pairwise comparisons of all other amplicons. At time zero there was no difference in amplification success of the four amplicons, but with increased ADD the probability of getting the 194, 305, and 384 bp amplicons was 4, 6.15, and 5.92 times lower than getting the 70 bp amplicons ($p < 0.001$ for all permutations).

© 2013 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 2. Shown below are the results of 4-plex multiplex amplification of pig soft muscle tissues project

Days	ADD	Pig whole carcasses			
		70 bp	194 bp	305 bp	384 bp
(A) February 2010					
14	34	+++	+++	+++	+++
21	57	+++	+++	+++	+++
29	88	+++	+++	+++	+++
41	185	+++	+++	+++	+++
57	302	+++	+++	+++	+++
64	367	+++	+++	+++	+++
71	418	+++	+++	+++	+++
77	494	+++	+++	+++	+++
84	554	+++	++−	+−−	+−−
90	603	+++	++−	+−−	+−−
(B) May 2010 project					
6	77	+++	+++	+++	+++
13	159	+++	+++	+++	+++
16	203	+++	+++	+++	+++
24	295	+++	+++	+++	+++
27	338	+++	+++	+−−	+−−
(C) September 2010					
9	106	+++	+++	+++	+++
11	137	+++	+++	+++	+++
15	210	+++	+++	+++	+++
18	254	+++	+++	+++	+++
22	288	+++	+++	+++	+++
30	342	+++	+++	+++	+++
38	420	+++	−−−	−−−	−−−
44	490	+++	+−−	−−−	−−−

"+" denotes the presence and "−" represents the absence of a PCR amplicon.

The probability of getting 305 and 384 bp was 1.54 and 1.58 times less than the 194 bp amplicons ($p < 0.001$ for both permutations).

4 Discussion

A multiplex has been designed that will amplify four amplicons between 70 and 384 bp, which is a size range typically analyzed using commercial kits for forensic DNA profiling. The primers work robustly using between 2.80 and 0.3 ng of template DNA from humans, pigs, and rabbits, which gives the multiplex more utility than one that is specific to human DNA. In addition, the addition of the IACs to the multiplex allows for the simultaneous detection of PCR inhibitors. Similar multiplexes have been described in the literature previously for carrying out persistence studies, but to our knowledge this is the first system described that will work with humans and other mammalian species that can be used as model systems. The system was designed, in part, to be used for taphonomic studies at the University of Central Lancashire's TRACES facility, where pigs and rabbits are used as experimental models [29, 36, 44]. Designing a multiplex that can also be used with human DNA does offer some advantages to making it species-specific. When using it in the field experiment described, an advantage was that commercial human DNA standards could be obtained—these are routinely supplied as components of commercial quantification kits and PCR amplification kits. This allowed accurate quantification standards to be used in real-time PCR and also gives more confidence when assessing the sensitivity of the assay than using only in-house standards.

The utility of the multiplex has been tested in a series of taphonomic experiments to assess the persistence of DNA in muscle tissues. The trends seen in the results are as expected, with warmer temperatures leading to higher rates of DNA degradation. What is somewhat surprising is the time that DNA is present in the muscle tissues—even in summer; with daytime temperatures frequently exceeding 20°C, DNA over 384 bp long could be detected up to 24 days. In May a factor limiting the retrieval of DNA was the availability of soft muscle tissue, which was not present after 27 days. In the February series, the muscle tissue maintained its mechanical integrity and did not change much in appearance; in the May and September experiments, the muscle tissue again maintained its physical properties until the very last stages of the experiment (typically the penultimate or final sample) where the mechanical integrity of the muscle broke down with the muscle becoming soft and sticky—once the muscle reached this state it did not persist for more than 3–4 days. In all studies we did not see significant changes in muscle color with only slight fluctuations between pink and red.

The persistence of DNA in muscle tissue could influence practices in sampling of cadavers for the purpose of human identification, at least in temperate climates, with DNA successfully extracted from muscle recovered from what would be considered by most to be highly decomposed bodies. The results presented here can be compared with other studies, but direct comparison is difficult. A similar experiment carried out in Australia presented unusual results in that DNA could not be recovered and amplified from samples on day 2, and in several cases on day 0 [3], which suggest that either the PCR assay or extractions were not robust. Other experiments used excised tissues rather than whole animals [5], and in some cases these were not exposed to the environment [8, 14], which will have affected the DNA persistence. It is necessary to carry out more detailed studies in warmer climates before making any generalized statements about how long DNA might be available in muscle tissue. When applying the conclusions of these experiments to forensic casework care also has to be taken not to overinterpret the results as these are based on an experimental animal model. However, pigs have been used in many taphonomic studies as they have several features, such as size, anatomy and metabolic rate similar to humans and have been used extensively as a model in taphonomic research in Europe [29, 44–46].

Another utility of the multiplex is that it can be used for studying factors that affect persistence in other types of forensic evidence. Several studies have been described that have used commercial PCR kits to assess the level of degradation

© 2013 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

[12–14]. This multiplex offers advantages over the commercial kits in that it is much less expensive and it is also amplifying homozygous nonvariable loci, removing the complication of samples from different donors having different sized alleles and combinations of homozygous and heterozygous loci. The addition of the IACs enhances the utility of this system over commercial systems in that the effect of inhibitors over the range of the multiplex can be assessed.

There is also the potential to use a multiplex such as this one as a screening tool for forensic analysis of material that is believed to be degraded, because such multiplex can potentially provide valuable information and minimize the waste of expensive reagents. Looking at the results in Table 2, it is possible to see samples where there is only amplification of the 70 and 194 bp amplicons, which would indicate that the use of a kit that is optimized to amplify small amplicons would be most beneficial. In other cases, the 70 bp amplicon had persisted after the other amplicons failed to be amplified—this would highlight the benefit of analyzing the samples with an alternative markers system that could utilize smaller amplicons, such as the SNPs and indels, where amplicons can be as small as 70 bp. The addition of internal amplified controls [30, 31] allows detection of PCR inhibitors and differentiation between DNA degradation and PCR inhibition, thereby enhancing the multiplex's ability to be used as a tool to assess degradation in different types of forensic evidence and to be used as a screening tool for complex forensic samples. However, if the system is used for screening complex samples to assess for inhibitors, it should be considered that different commercial PCR kits have different sensitivities, for example the Minifiler kit is more robust to inhibitors than the Identifiler kit [19] and in general later generations of commercial kits have higher resistance to inhibitors. Therefore, the relative impact of inhibition on this multiplex and the commercial kit of choice would have to be assessed to avoid overinterpretation. It should be noted that this same limitation exists with real-time PCR detection of inhibition in commercial kits, where the sensitivity of the detection varies between the real-time systems, for example Quantifiler and PCR multiplexes [30]. The best approach would be to incorporate the IACs into the commercial multiplex kit, however, this is a complex process [30, 31] and may not be practical in a commercial environment.

We would like to express our thanks to the staff and students working at TRACES, University of Central Lancashire, UK. In particular, we would like to thank Tal Simmons and Peter Cross for providing access to the facilities and logistical support throughout the project.

The authors have declared no conflict of interest.

5 References

[1] Johnson, L. A., Ferris, J. A. J., *Forensic Sci. Int.* 2002, *126*, 43–47.

[2] Fredericks, J. D., Bennett, P., Williams, A., Rogers, K. D., *Forensic Sci. Int.* 2012, *6*, 375–380.

[3] Larkin, B., Iaschi, S., Dadour, I., Tay, G. K., *Forensic Sci. Med. Pathol.* 2010, *6*, 83–92.

[4] Kaiser, C., Bachmeier, B., Conrad, C., Nerlich, A., Bratzke, H., Eisenmenger, W., Peschel, O., *Forensic Sci. Int.* 2008, *177*, 32–36.

[5] Phengon, P., Wongwiggarn, S., Panvisavas, N., *Forensic Sci. Int.* 2008, *1*, 439–441.

[6] Swango, K. L., Hudlow, W. R., Timken, M. D., Buoncristiani, M. R., *Forensic Sci. Int.* 2007, *170*, 35–45.

[7] Foran, D. R., *J. Forensic Sci.* 2006, *51*, 766–770.

[8] Itani, M., Yamamoto, Y., Doi, Y., Miyaishi, S., *Acta Med. Okayama* 2011, *65*, 299–306.

[9] Alaeddini, R., Walsh, S. J., Abbas, A., *Aust. J. Forensic Sci.* 2010, *42*, 211–220.

[10] Bar, W., Kratzer, A., Machler, M., Schmid, W., *Forensic Sci. Int.* 1988, *39*, 59–70.

[11] Imaizumi, K., Miyasaka, S., Yoshino, M., *Sci. Justice* 2004, *44*, 199–208.

[12] Dixon, L. A., Dobbins, A. E., Pulker, H. K., Butler, J. M., Vallone, P. M., Coble, M. D., Parson, W., Berger, B., Grubwieser, P., Mogensen, H. S., Morling, N., Nielsen, K., Sanchez, J. J., Petkovski, E., Carracedo, A., Sanchez-Diz, P., Ramos-Luis, E., Brion, M. I., Irwin, J. A., Just, R. S., Loreille, O., Parsons, T. J., Syndercombe-Court, D., Schmitter, H., Stradmann-Bellinghausen, B., Bender, K., Gill, P., *Forensic Sci. Int.* 2006, *164*, 33–44.

[13] Rubio, L., Martinez, L. J., Martinez, E., Martin de las Heras, S., *J. Forensic Sci.* 2009, *54*, 1411–1413.

[14] Niemcunowicz-Janica, A., Pepinski, W., Janica, J. R., Skawronska, M., Janica, J., Koc-Zorawska, E., *J. Forensic Sci.* 2007, *52*, 867–869.

[15] Alaeddini, R., Walsh, S. J., Abbas, A., *Forensic Sci. Int.* 2010, *4*, 148–157.

[16] Interpol. 2009. Disaster Victim Identification Guide. www.interpol.int/Media/Files/INTERPOL-Expertise/DVI/DVI-Guide.

[17] Prinz, M., Carracedo, A., Mayr, W. R., Morling, N., Parsons, T. J., Sajantila, A., Scheithauer, R., Schmitter, H., Schneider, P. M., *Forensic Sci. Int.* 2007, *1*, 3–12.

[18] Butler, J. M., Shen, Y., McCord, B. R., *J. Forensic Sci.* 2003, *48*, 1054–1064.

[19] Coble, M. D., Butler, J. M., *J. Forensic Sci.* 2005, *50*, 43–53.

[20] Alenizi, M. A., Goodwin, W., Hadi, S., Alenizi, H. H., Altamar, K. A., Alsikel, M. S., *J. Forensic Sci.* 2009, *54*, 350–352.

[21] Freire-Aradas, A., Fondevila, M., Kriegel, A.-K., Phillips, C., Gill, P., Prieto, L., Schneider, P. M., Carracedo, A., Lareu, M. V., *Forensic Sci. Int.* 2012, *6*, 341–349.

[22] Phillips, C., Fang, R., Ballard, D., Fondevila, M., Harrison, C., Hyland, F., Musgrave-Brown, E., Proff, C., Ramos-Luis, E., Sobrino, B., Carracedo, A., Furtado, M. R., Syndercombe-Court, D., Schneider, P. M., *Forensic Sci. Int.* 2007, *1*, 180–185.

[23] Sanchez, J. J., Phillips, C., Børsting, C., Balogh, K., Bogus, M., Fondevila, M., Harrison, C. D., Musgrave-Brown, E., Salas, A., Syndercombe-Court, D., Schneider, P. M.,

© 2013 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Carracedo, A., Morling, N., *Electrophoresis* 2006, *27*, 1713–1724.

[24] Kidd, K. K., Pakstis, A. J., Speed, W. C., Grigorenko, E. L., Kajuna, S. L. B., Karoma, N. J., Kungulilo, S., Kim, J.-J., Lu, R.-B., Odunsi, A., Okonofua, F., Parnas, J., Schulz, L. O., Zhukova, O. V., Kidd, J. R., *Forensic Sci. Int.* 2006, *164*, 20–32.

[25] Pereira, R., Phillips, C., Alves, C., Amorim, A., Carracedo, A., Gusmao, L., *Electrophoresis* 2009, *30*, 3682–3690.

[26] 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., *Forensic Sci. Int.* 2012, *6*, 469–476.

[27] Vass, A. A., Barshick, S. A., Sega, G., Caton, J., Skeen, J. T., Love, J. C., Synstelien, J. A., *J. Forensic Sci.* 2002, *47*, 542–553.

[28] Megyesi, M. S., Nawrocki, S. P., Haskell, N. H., *J. Forensic Sci.* 2005, *50*, 618–626.

[29] Simmons, T., Adlam, R. E., Moffatt, C., *J. Forensic Sci.* 2010, *55*, 8–13.

[30] Zahra, N., Hadi, S., Smith, J. A., Iyengar, A., Goodwin, W., *Electrophoresis* 2011, *32*, 1371–1378.

[31] Zahra, N., Hadi, S., Goodwin, W., *Electrophoresis* 2012, *33*, 2833–2839.

[32] Rozen, S., Skaletsky, H., *Methods Mol. Biol.* 2000, *132*, 365–86.

[33] Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., Leunissen, J. A. M., *Nucleic Acids Res.* 2007, *35*, W71–W74.

[34] Kibbe, W. A., *Nucleic Acids Res.* 2007, *35*, W43–W46.

[35] Swango, K. L., Timken, M. D., Chong, M. D., Buoncristiani, M. R., *Forensic Sci. Int.* 2006, *158*, 14–26.

[36] Cross, P., Simmons, T., Cunliffe, R., Chatfield, L., *Forensic Sci. Policy Manage.* 2010, *1*, 187–191.

[37] Nazir, M. S., PhD Thesis, University of Central Lancashire, Lancashire 2011.

[38] Bernstein, R. M., Schluter, S. F., Bernstein, H., Marchalonis, J. J., *Proc. Natl. Acad. Sci. USA* 1996, *93*, 9454–9459.

[39] Bernstein, R. M., Schluter, S. F., Lake, D. F., Marchalonis, J. J., *Biochem. Biophys. Res. Commun.* 1994, *205*, 687–692.

[40] Groth, J. G., Barrowclough, G. F., *Mol. Phylogenet. Evol.* 1999, *12*, 115–123.

[41] Gill, P., Buckleton, J., *Forensic Sci. Int.* 2010, *4*, 221–227.

[42] Gill, P., Whitaker, J., Flaxman, C., Brown, N., Buckleton, J., *Forensic Sci. Int.* 2000, *112*, 17–40.

[43] Smith, J. A., Baker, N. C., *Forensic Sci. Int.* 2008, *1*, 620–622.

[44] Cross, P., Simmons, T., *J. Forensic Sci.* 2010, *55*, 295–301.

[45] Turner, B., Wiltshire, P., *Forensic Sci. Int.* 1999, *101*, 113–122.

[46] Dekeirsschieter, J., Verheggen, F. J., Gohy, M., Hubrecht, F., Bourguignon, L., Lognay, G., Haubruge, E., *Forensic Sci. Int.* 2009, *189*, 46–53.

© 2013 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim