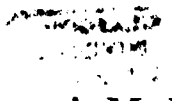
The Analysis of Human Mitochondrial DNA in Peninsular Malaysia

Thesis submitted in accordance with the requirements of the University of Glasgow for the Degree of Doctor of Philosophy

By

Zafarina Zainuddin
B.Sc. (Hons), M.Sc.



Department of Forensic Medicine and Science
University of Glasgow
May 2004

ACKNOWLEDGEMENT

First of all, I would like to thank my supervisor, Dr William Goodwin, for all the efforts, support and guidance he had provided along my study. His interest and commitment in this research had given me the drive and courage to complete this work. I would also like to take this opportunity to express my gratitude to my sponsors, University Science of Malaysia and Malaysian Department of Public Services. Many thanks to the Malaysian Department of Orang Asli Affairs for helping with the samples collection. Special thanks to the Department of Forensic Medicine and Science and the Division of Earth Sciences for providing me all the facilities needed for this study. Also special thanks to Dr John Oliver, my co-supervisor and all staffs in the department, especially Alison, Marie, Alan and Carol Anne for their helping hand.

I would also like to thank Dr Lee Chatfield and all staffs especially Dr Michelle Whittaker and Dr Syed Sibte Hadi, in the Department of Forensic and Investigative Science, University of Central Lancashire, for the facilities and help, which make the completion of this study, become possible. Also thank you to Dr Maggie Cusack, Sandra Tierney and John Harley for their technical advice and being very helpful during my days in the molecular palaeontology lab. Many thanks to Georgia and Julie in the MBSU unit for all the sequencing work.

Million of thanks to all my friends, especially to Afaf, Nina, Aida, Farah, Dikla and Mitra for always being there for me. Very special thanks to my parents, Hj Zainuddin and Hjh Jallahah, who had sacrificed a lot all this while.

Finally, I would like to dedicate this piece of work to my beloved husband, Muhsin, who had never complained for all the lonely minutes and hours I've made him gone thru during my busy days. This work is also dedicated to my children, Fitri and Aina and also to all my family members. Thank you for all the love that keeps me going.

LIST OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

bp	base pair
CEPH	Centre d'Etude Polymorphisme Human
DNA	Deoxyribonucleic acid
dROX™	6-carboxy-X-rhodamine
ddNTPs	Dideoxynucleotide triphosphates
dTAMRA™	N,N,N′,N′-tetramethyl-6-carboxyrhodamine
EDTA	Ethylenediaminetetraacetic acid
EtBr	Ethidium bromide
Exo I	*Exonuclease* I
HLA	Human Leucocyte Antigen
HRP-SA	Horseradish peroxidase-streptavidin
HVI	Hypervariable region I
HVII	Hypervariable region II
FAM	Fluorescein
MLP	Multi Locus Probe
mtDNA	Mitochondrial DNA
NaOH	Sodium hydroxide
PCR	Polymerase Chain Reaction
TMB	3,3′,5,5′-tetramethylbenzidine
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribonucleic Acid
SAP	Shrimp Alkaline Phosphatase
SDS	Sodium dodecyl sulfate
SINEs	Short Interspersed Repeats
SLP	Single Locus Probe
SNP	Single Nucleotide Polymorphism
STRs	Short Tandem Repeats
tRNA	Transfer Ribonucleic Acid
VNTRs	Variable Number of Tandem Repeats
YBP	Years Before Present

ABSTRACT

Mitochondrial DNA analysis was undertaken on samples collected from two populations in Peninsular Malaysia, the Modern Malay (102 samples) and Orang Asli (59 samples from Jahai and Kinsiu sub-groups). The hypervariable region I (HVI) of the mtDNA control region was amplified and sequenced. Polymorphisms were reported by aligning each sequence to the Cambridge Reference Sequence (CRS). A total of 94 polymorphisms were observed in the Modern Malay samples, which formed 75 different haplotypes. The Orang Asli showed notably lower number of the HVI region variations, with only 28 polymorphisms and 13 haplotypes observed. Genetic diversity calculated for the Modern Malays and Orang Asli were 0.989 and 0.818, respectively. Probability of random match calculated was 0.0202 for the Modern Malays and 0.1962 for the Orang Asli.

The mtDNA coding region variations was examined using RFLP analysis. Combination of both RFLP and HVI sequence data had placed the Modern Malays into three major Southeast Asian haplogroups, M, B and F. These findings had initially suggested that the Modern Malays shared a common lineage with other populations within this region. Two novel sub-clusters, M21a and R21 were found at a high frequency within the Orang Asli samples. These sub-clusters, which have also been found in other Semang sub-groups appear to be indigenous Semang haplogroups. The limited number of mtDNA haplotypes shared between the Modern Malays and Orang Asli suggested a discontinuity of mtDNA between these populations. Even though both populations were believed to be among the earliest populations of Peninsular Malaysia, this result indicates that the Modern Malays were not direct descendants of the Orang Asli.

Minisequencing analysis was carried for further interrogation of the mtDNA coding region polymorphisms. Single nucleotide polymorphisms, which were responsible for creating RFLP sites used in haplogroups determination, were examined using SNaPshot™ Multiplex kit from ABI PRISM®. A total of 17 SNPs were examined in two multiplex sets, which contained 10 and 7 minisequencing primers each. This method was proven to be a robust and efficient way for SNP screening, with the ability of multiplexing and highly automated analysis. Sequencing of the whole PCR product and RFLP analysis were found to be unnecessary in future since minisequencing could provide the same qualitative characteristics and results from both methods, with less time and effort consumed.

Besides mtDNA analysis, the autosomal STR markers were also examined using PowerPlex® 16 system for both populations. These data could provide more information when added to the available STR database for Malaysian populations.

1. INTRODUCTION

1.1 Forensic DNA Analysis

Rapid development of DNA technology in the past 15 years had increased the ability of human identification in forensic science. The introduction of DNA fingerprinting by an English scientist, Sir Alec Jeffreys in 1985 (Jeffreys *et al.*, 1985) has had an enormous impact in forensic science. Before the use of molecular biology became available, identification of biological evidence mostly depended upon the analysis of classical fingerprint and blood group marker systems. However these methods have a number of limitations such as limited source material. Conventional blood typing methods were dependent upon the availability of blood or body fluid from the crime scene whilst DNA profiling could be performed on any nucleated biological material such as hair, nail, saliva and seminal stains. Results of forensic DNA profiling are now widely accepted in courts as important evidence in crime cases and also a powerful tool in human identification.

Human DNA is basically composed of the 'coding' and 'non-coding' regions. The 'coding' region only makes up about 3% of human genomic DNA. The remaining 97% does not code for genetic information and in most cases does not undergo the same selection pressures as coding regions. The mutations are usually not fatal and can be transmitted to the offspring. Over time this has resulted in tremendous genetic variability (Schneider, 1997). DNA markers used for human identity testing are taking the advantage of this special feature. Several DNA typing technologies, as well as different DNA markers have been established for use in forensic DNA analysis. Each

technique and marker systems serves different degree of usefulness in different forensic applications (Table 1.1).

1.1.1 Restriction Fragment Length Polymorphism (RFLP)

The first DNA analysis method used in forensic science was restriction fragment length polymorphism (RFLP). Sir Alec Jeffreys had discovered a stretch of repeated DNA sequences surrounding the chromosomal centromere. This region typically consists of 500 to 10,000 base pairs, comprising tandemly repeated units, each unit being 15 to 35 base pairs in length and known as minisatellites or variable number of tandem repeats (VNTRs). The number of repeated unit and length of the region could differ significantly between individuals. Analysis was carried out by digestion of DNA with a restriction endonuclease followed by Southern blotting and hybridisation with a specific probe. Many loci have been established for RFLP analysis, which increase the chance of finding differences between two people.

RFLP techniques offer a high power of discrimination when using multi-locus probes (MLP). When undertaking MLP analysis, southern blotting was carried out at low stringency conditions. The radio-labelled probes hybridised with a series of tandemly repeated DNA sequences to produce a 'genetic fingerprint'. The resulting autoradigraphs display a unique ladder like pattern, which looked similar to a supermarket bar code. However this method required relatively large amounts of high molecular weight DNA to be extracted from the crime scene samples and a great deal of labour, time and expertise to produce a DNA profile. Since most evidence left at the crime scene were just traces and sometimes severely degraded, this technique often

failed to produce reliable results. MLP profiles are also difficult to interpret which was one of their major downfalls.

The introduction of single locus probe (SLP) had simplified the interpretation in RFLP analysis. In this new system, the probes identified repeated sequences in individual minisatellites using high stringency conditions during Southern blotting to produce a maximum of two bands per probe (Wong *et al.*, 1987). A number of loci were sequentially analysed by electrophoretic separation along with a control ladder of known molecular size fragments in order to obtain an estimate size of the minisatellites bands. Since the result could be recorded as a numerical set, the interpretation of SLP systems was much easier and clearer than MLP analysis.

Table 1.1: List of DNA typing technologies and markers used in forensic DNA analysis and the advantages and disadvantages of each technique/marker.

RFLPs: Restriction Fragment Length Polymorphisms

VNTRs: Variable Number of Tandem Repeats

STRs: Short Tandem Repeats

PCR: Polymerase Chain Reaction

mtDNA: Mitochondrial DNA

SNPs: Single Nucleotide Polymorphisms

SINEs: Short Interspersed Nuclear Elements

MLP: Multi Locus Probe

SLP: Single Locus Probe

DNA typing technology/ Markers	Advantages	Disadvantages
RFLPs/		
VNTRs: MLP SLP	-High discrimination power -Highly variable between individuals	-Requires high amount of HMW DNA (10 to 50 ng) -Time consuming -Intensive labour/expertise
PCR amplification/		
HLA DQα Polymarker	-smaller size, easy amplification -rapid analysis -small amount of template (0.2 to 0.5 ng)	-Single locus analysed, lower power of discrimination than RFLP
Autosomal STRs	-High discrimination power -Highly variable between individuals -Easily amplified by PCR due to the small alleles (~100 bp to ~400 bp). -Permits multiplexing	-Failed to produce result from severely degraded/damaged DNA -Stutter artefacts -expensive multiplex kit
Y STRs	-Paternal inheritance mode -Useful to identify male component in mixed sample -Tracing human evolution through male lineage	-Not highly variable between individuals -Many markers needed to obtain high degree of discrimination
mtDNA	-Maternal inheritance mode -Higher copies per cell than nuclear DNA -Efficiently amplified from severely degraded or damaged biological samples -Tracing human evolution through female lineage -Heteroplasmy characteristic	-Low power of discrimination -Treated as single locus -Need large database for better random match probability statistic.
SNPs	-High occurrence in human genome, about once every 1 kb -Can be multiplex up to thousand markers on microchip -No stutter artefacts -Size based separation not needed -Smaller amplicons (~100 bp), higher chance to amplify from degraded DNA samples	-Less informative than STRs -Biallelic marker, more loci needed to get a high discrimination power
SINEs/ *Alu* insertion	-Stable genetic marker -Easily genotype by PCR and agarose gel (ethidium bromide stained) electrophoresis	-Exhibit less variation than STRs

1.1.2 Polymerase Chain Reaction (PCR)

Many forensic DNA samples were not at sufficient quantity and/or quality for them to be analysed using VNTR technology. The innovation of PCR technique by Kary Mullis in 1986 (Mullis *et al.*, 1986) had opened a new era in DNA profiling. The ability of PCR technology to amplify from minute quantities and also degraded DNA templates had enhanced the usefulness of DNA profiling in forensic science. PCR is an *in vitro* enzymatic process, which amplifies a specific region of DNA over and over again to yield million copies of the target sequence (Saiki *et al.*, 1988). The advance technique of multiplex PCR reactions allow many target sequences to be amplified simultaneously, adding additional advantage to the PCR technique, which is fast, sensitive and reliable for forensic DNA analysis.

The first system for forensic analysis of PCR-amplified DNA was the HLA DQα system (Perkin-Elmer). This locus showed some differences in the DNA sequences between different individuals and the variation is detected by specific probes targeted to particular sub regions within the locus. These probes detected six common DQα alleles, which in combination determine 21 possible genotypes. Since only one locus is being analysed in this system, the power of discrimination is not as high as VNTR. The HLA DQα was soon replaced by STR system, which had higher power of discrimination.

1.1.3 Short Tandem Repeats (STRs)

Human DNA was found to have tandemly repeated DNA sequences which varied in size. The region with repeated units between 2 to 6 bp in length were called

microsatellites or short tandem repeats (STRs). STR markers are scattered along human DNA and appeared on average of every 10,000 nucleotides (Edwards *et al.*, 1991). The numbers of nucleotides for each repeat unit within an STR locus were normally the same but the number of repeat units for each locus may differ, resulting many different lengths of alleles. These variations had made STR loci very polymorphic and therefore well suited for human identification purposes. STR markers were usually identified by searching DNA sequence databases for region with more than six contiguous repeat units (Weber and May, 1989) or by molecular biology isolation method (Edwards *et al.*, 1991).

With the avaibility of PCR, STR became a popular DNA marker since they were easy to amplify and highly variable between individuals. PCR amplification of degraded DNA, which is very common in forensic sample, is more successful for smaller size products. Since the STR alleles size is small (~100 to 400 bp) compared to VNTR (~400 bp to >20 kb), it is a better and promising marker for forensic DNA profiling. The problem of preferential amplifications and allele dropout can also be reduced in heterozygous individuals when using smaller DNA repeats marker since both alleles are similar in size. The region of interest containing repeat sequences was simply amplified using a set of primers complimentary to the flanking regions of the target sequences.

The majority of STRs being used in the forensic community consist of four-nucleotide repeat units, also known as tetranucleotide STRs (Kimpton *et al.*, 1993; Urquhart *et al.*, 1995). The amplified STR loci are analysed by electrophoresis to separate the alleles according to size. Popular detection methods include the silver staining, which is low in cost and fluorescent techniques, which can be highly automated.

One of the first STR multiplex systems, the quad, contained 4 loci (THO1, FES/FPS, VWA and F13A1) by the Forensic Science Service (Kimpton *et al.*, 1994), followed by the second-generation multiplex (SGM) made up of 6 loci (TH01, VWA, FGA, D8S1179, D18S51 and D21S11) (Gill *et al.*, 1996; Sparkes *et al.*, 1996). To date, several STR kits are available in the market, which allows robust multiplex amplification up to 16 STR markers. With match probability exceeding one in a billion, amplification with 1 ng or less DNA template and results available within just few hours, these STR kits has truly revolutionised forensic DNA.

1.1.4 Y-chromosome markers

In some forensic situations, especially in sexual assaults cases, the ability to identify male component from a mixture of samples is very valuable. Y-chromosome markers were also used in tracing human evolution through male lineages (Jobling and Tyler-Smith, 1995), paternity testing and tracing paternal lineages, as demonstrated in the Thomas Jefferson and Eston Hemmings case (Foster *et al.*, 1998). Within the past several years, a number of Y-chromosome markers had been developed. These markers fall into two broad categories (Jobling and Tyler-Smith, 2000); i) Biallelic markers with low mutation rate-examples are single nucleotide polymorphisms and *Alu* element insertion; ii) Multiallelic markers with higher mutation rates-examples are microsatellites and minisatellites. Combination of both biallelic and multiallelic polymorphisms had succesfully used in the analysis of population structuring (Ciminelli *et al.*, 1995; Scozzari *et al.*, 1997) and tracking paternal lineages (Foster *et al.*, 1998).

1.2 Human Mitochondrial DNA

Mammalian cells possess two different types and interdependent genomes, the nuclear genome and mitochondrial genome. Mitochondria are semi-autonomously functioning organelles containing a resident genome that undergoes replication, translation and transcription of their own DNA (Anderson *et al.*, 1981, Taanman, 1999). The main function of these organelles is to generate energy for the cells. They produce almost 90% of the cellular ATP through oxidative phosphorylation (Wallace, 1997). Each organelle in most mammalian cells contains 2-10 copies of mitochondria DNA (mtDNA) molecules and each cell may contain more than 1,000 copies of mtDNA (Wei, 1992). Mitochondrial DNA is also well known for exhibiting maternal inheritance. All mtDNA molecules in sperm degraded shortly after fertilisation, leaving only mtDNA from the mother in all cells of the progeny. Barring mutation, all maternally related individuals should have the same mtDNA profile. The complete sequence of the human mtDNA was first obtained in 1981 (Anderson *et al.*, 1981). The revise version, where few corrections have been made, is known as Cambridge Reference Sequence (CRS) (Andrews *et al.*, 1999). This sequence is used as comparison to report polymorphisms from new mtDNA sequences.

1.2.1 Organisation of mtDNA

Human mtDNA is a supercoiled double stranded circular molecule, comprised of purine-rich strand and pyrimidine-rich strand which commonly called as the heavy and light strands, respectively. It bears 37 genes coding for 13 polypeptides of mitochondrial electron respiratory chain, 22 tRNAs and 2 rRNAs needed for the

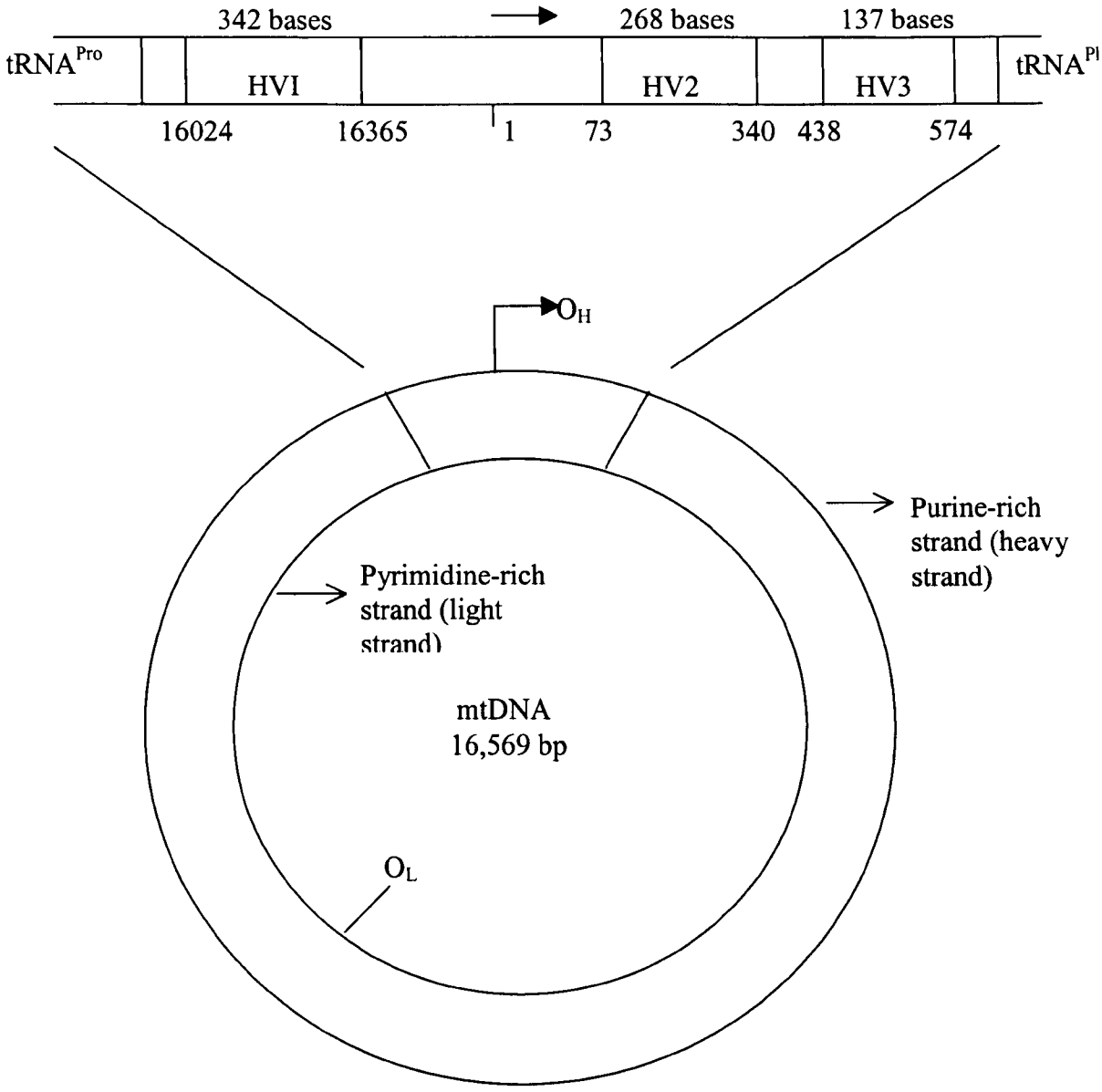
synthesis of polypeptides (Penta *et al.*, 2001). Human mtDNA is approximately 16.5 kb in length and among the smallest mitochondrial genomes found in eukaryotes (Penta *et al.*, 2001). mtDNA replicates using two different origin of replications, each situated at the heavy and light strand. Nucleotide positions in mtDNA genome are numbered according to the Anderson Reference Sequence (Anderson *et al.*, 1981). The numerical designation begins at the origin of replication of the heavy strand and continues around the circle for approximately 16,569 base pairs (Wilson *et al.*, 1993).

The control region which includes the displacement-loop (D-loop) is the most rapidly evolving part of mtDNA (Upholt and David 1977; Carracedo *et al.*, 2000;Herrnstadt *et al.*, 2002). This non-coding region is approximately 1,100 bp in length (Figure 1.1), situated between the mitochondrial $tRNA^{pro}$ and $tRNA^{phe}$ genes. It contained the origin of replication for the heavy strand synthesis, both mitochondrial transcription promoters and also served as the core site for mtDNA replication and transcription (Taanman, 1999; Lutz, *et al.*, 1997). The evolutionary rate of this region is almost 10 times more than the gene coding region and therefore variation is very much concentrated in this region (Parsons and Coble, 2001). The variable sites in the control region consist of three hypervariable segments. Hypervariable region I (HVI) ranges from position 16,024 to 16,365, hypervariable region II (HVII) extends from position 73-340 (Greenberg *et al.*, 1983; Wilson *et al.*, 1993) and hypervariable III situated between position 438 to 574 (Lutz *et al.*, 1997). These boundaries are however not rigidly defined and vary between studies or laboratories. Both HVI and HVII regions are well known as hotspots for base pair substitutions (Wallace *et al.*, 1995)

Figure 1.1: Schematic diagram of human mitochondrial DNA genome, showing the location of HVI, HVII and HVIII regions within the control region.

(Edited from Butler, J.M. 2001; Garritsen *et al.*, 2001)

Figure showing simplified diagram of the circular mitochondrial genome, emphasising the position of the hypervariable regions. D-loop region is bound by the genes for tRNA proline ($tRNA^{Pro}$) and tRNA phenylalanine ($tRNA^{Phe}$). The purine-rich strand and pyrimidine-rich strand were also called as the heavy and light strand, respectively.



O_H : Origin of replication of heavy strand

O_L : Origin of replication of light strand

→ : Showing the direction of nucleotide numbering in mtDNA

1.2.2 Advantanges of mtDNA in Forensic DNA Analysis

Over the years, there are hundreds of unidentified remains which are difficult and almost impossible to be identified using conventional methods such as dental comparison, fingerprint analysis or facial reconstruction. DNA analysis, in particular mtDNA analysis, has played a key role in identification of these remains. The most notable case, which displays the utility, strength and reliability of mtDNA analysis, was the identification of the Romanov family (Gill *et al.*, 1994). Since the genetic elements of mtDNA survive without variation by genetic recombination, mtDNA sequences should be identical to the maternally related individuals, unless a mutation occurred. This maternal inheritance feature of mtDNA is very useful in tracking families and population, as well as identification of individuals involved in mass disaster, missing person and criminal cases. In the case of the Romanov family, a perfect match of mtDNA sequence was found between the putative bones of Tsarina and her three daughters with the maternally related Prince Philip of United Kingdom (Gill *et al.*, 1994).

Mitochondrial DNA is present as many as 50,000 molecules in a single oocyte (Bender *et al.*, 2000) to an average of 5000 molecules in epithelial cells (Bogenhagen and Clayton, 1974). The high copy number of mtDNA compared to only one copy of nuclear DNA in each cell has increased the possibility for it to be recovered from limited biological samples or even from severely degraded samples. Hitherto, mtDNA had been successfully extracted from teeth (Boles *et al.*, 1995; Ginther *et al.*, 1992), hair shafts (Hopgood *et al.*, 1992; Wilson *et al.*, 1995b), bone fragments (Sullivan *et al.*,

1992; Kurosaki *et al.*, 1993) and even human faeces (Hopwood *et al.*, 1996), in addition to the normally used samples such as blood, semen and saliva.

The forensic value of mtDNA is even better with the sequence variation of the non-coding region. Two hypervariable segments (HVI and HVII) within the non-coding region, tend to mutate 5 to 10 times faster than nuclear DNA (Morley *et al.*, 1999). This region has been estimated to vary about 1-3% between unrelated individuals (Piercy *et al.*, 1993). The high mutation rate makes this region as a useful tool in human identity testing (Holland and Parsons, 1999). In addition, mtDNA had a low efficiency of repair and high frequency of replication error particularly in the non-coding region (Wilson *et al.*, 1993), making it highly polymorphic and suitable as a genetic marker. Sequence data of the control region had been characterized from various populations worldwide. Studies of human mtDNA variation had also played an important role in understanding the evolutionary origin and population genetic history of modern *Homo sapiens* (Holland and Parsons, 1999).

1.2.3 Heteroplasmy

Human mtDNA typically existed as monoclonal within an individual but a condition called heteroplasmy does occur (Comas *et al.*, 1995). This phenomenon is defined as locating more than one mtDNA type within an individual. Heteroplasmy could be observed in several forms: i) more than one mtDNA type in a single tissue ii) different mtDNA types in different tissues and iii) homoplasmic in one tissue and heteroplasmic in another tissue (Budowle *et al.*, 1999; Stewart *et al.*, 2001). Both length and sequence heteroplasmy occur. In sequence heteroplasmy, more than one base is found at a

particular site of mtDNA sequence. Heteroplasmy at two or even three sites may also occur, but at much lower rate (Budowle *et al.*, 2002). The most frequent length heteroplasmy involved a C-stretch region, where a T to C transition at position 16,189 resulted a stretch of 10 to 13 (or more) cytosines on the light strand of mtDNA molecules (Bendall and Sykes, 1995). In some cases, heteroplasmy can actually complicate the data interpretation of mtDNA analysis. A report had shown sequences from 12 hair shafts and saliva sample from one individual differing at position 16,093 (Sullivan *et al.*, 1996). The saliva sample was homoplasmic at that particular site while three of 12 hair shafts were heteroplasmic (C/T). If this was to occur in crime cases, it will lead to false exclusion.

In the other hand, heteroplasmy can also improve the probability of a match. The first sequence heteroplasmy accounted in the forensic identification involved the identification of the Russian Tsar Nicholas II (Gill *et al.*, 1994; Ivanov *et al.*, 1996). The mtDNA sequence of the Tsar Nicholas II putative remains was found to match the sequence of his living maternal descent except at one particular sequence. Tsar Nicholas II was found to have both T and C nucleotides at position 16,169. In order to confirm the remains were belonged to Tsar Nicholas II, his late brother Grand Duke of Russia Georgij Romanov was exhumed and tested with the same procedure (Ivanov *et al.*, 1996). His mtDNA sequence was found to be identical with Tsar Nicholas II, including bearing the same heteroplasmy at position 16,169. Since the extreme rarity of this heteroplasmy to happen by chance between unrelated individuals, the remains was declared as Tsar Nicholas II and put to rest with respect (Ivanov *et al.*, 1996; Butler, 2001).

1.3 Mitochondrial DNA Analysis

Several laboratories including the FBI Laboratory (Washington DC, USA), the Armed Forces DNA Identification Laboratory (Rockville, USA) and Forensic Science Service (Birmingham, UK) are currently using mtDNA in cases where nuclear DNA is too degraded or not in sufficient quantities to be typed (Butler and Levin, 1998). A number of methods have been developed for analysis of mtDNA and they are widely used either in forensic cases or population studies. The forensic application of mtDNA mostly involved the examination of control region since variations are very much concentrated in this region while the human population studies had occupied mtDNA polymorphisms from both coding and non-coding regions.

1.3.1 RFLP Analysis of the Coding Region

Restriction Fragment Length Polymorphism (RFLP) studies of mtDNA coding region had revealed a number of stable polymorphic site, many of which can be used to classify mtDNA haplogroups. It had been a powerful tool for detecting genetic variants between major populations. Low resolution restriction analysis involved the use of 5 to 6 restriction enzymes (Chen *et al.*, 2000) while high resolution analysis had occupied as much as 12 to 14 restriction enzymes (Cann *et al.*, 1987; Macaulay *et al.*, 1999). The analysis initiated by amplification of fragments containing the target mutation site within the coding region. These fragments were then digested using a set of restriction enzymes. The resulting restriction fragments were resolved by agarose or polyacrylamide gel electrophoresis and visualised by ethidium bromide/fluorescent dyes staining. The pattern of bands observed should be an indication either a mutation had

occurred or not, which lead to a site gain or loss. The most common set of restriction endonucleases used in the RFLP analysis includes the following enzymes: *Alu* I, *Ava* II, *Bam*H I, *Dde* I, *Hae* II, *Hae* III, *Hha* I, *Hinc* II, *Hinf* I *Hpa* I, Mse I, *Msp* I, *Mbo* I, *Rsa* I and *Taq* I.

Previous studies using high resolution RFLP analysis had successfully determined specific haplogroups for specific populations (Torroni *et al.*, 1994; Torroni *et al.*, 1996). Somehow, analysis using the common set of enzymes had only allowed examination of about 20% of the mitochondrial DNA genome (Wallace, 1994). Characterisation of mtDNA haplogroups was always helped by the sequence analysis of the HVI region. A good correlation has been obtained between RFLP data and HVI sequence data (Torroni *et al.*, 1996, 1998; Macaulay *et al.*, 1999), even with the problems of parallel mutations and back mutations of the hypervariable sites.

1.3.2 Direct Sequencing of Control Region

Application of mtDNA in forensic identification had taken place by using PCR to amplify several hundred base pairs in both HVI and HVII regions (Wilson *et al.*, 1995b). Primers used in the PCR usually amplified fragments from position 16,000 to 16,430 and 40 to 440 for HVI and HVII region, respectively. Since D-loop region only exhibit point mutations, examination of the amplified products was carried by direct sequencing. The fluorescent automated DNA sequencing, which is widely used these days, was first introduced in 1986 (Smith *et al.*, 1986). The reaction involved incorporation of fluorescently labelled ddNTPs into DNA fragments, which is later electrophoretically separated and detected by laser-induced fluorescence (Wilson *et al.*,

1993). In most studies, the amplified fragments were sequenced for both light and heavy strands in order to confirm that the information obtained is correct. Sequences obtained are aligned with a standard reference sequence, known as Anderson Reference Sequence (Anderson *et al.*, 1981), taking the name of the first author to publish the entire human mitochondrial genome sequences. Differences noted between a new sequence and the reference sequence were address as polymorphisms. Substitutions, deletion and insertion are different forms of polymorphisms observed in human mtDNA.

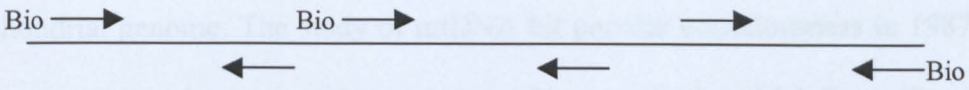
Frequency of each mtDNA type obtained from the sequencing result was defined by the number of occurrence of the particular sequence in the whole database (Piercy *et al.*, 1993). Even though sequencing is time consuming and relatively an expensive method, it is highly efficient and had so far generate large stock of established mtDNA control region data and provide a high degree of information for discriminating between unrelated individuals. Sequence analysis of HVI region had also been used to help in the haplogroup characterisation in population studies. To date, compilation of mtDNA data include thousands of sequences from HVI and HVII region from European, African, American and Asian populations (Budowle *et al.*, 1999; Wittig *et al.*, 2000; Imaizumi *et al.*, 2002; Fucharoen *et al.*, 2001).

1.3.3 Minisequencing

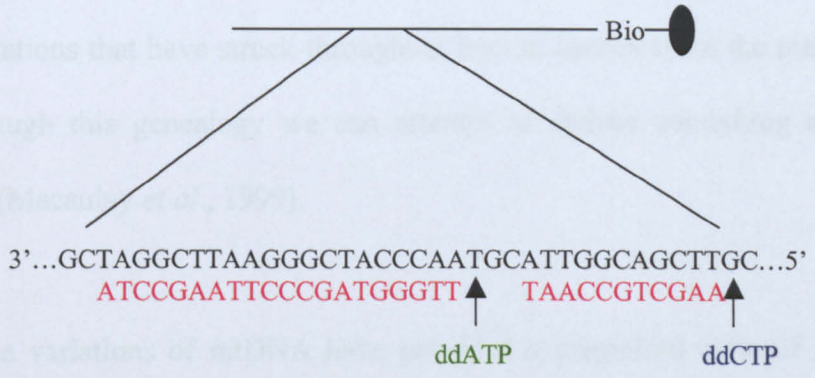
A rapid screening method, which can accelerate examination of large sample number, has a high forensic potential. Dealing with cases such as mass disaster, plane crash or identifying skeletal remains from mass graves would be a huge challenge for the forensic community if they were not facilitated with a robust and efficient analysis technique. Where typing on mtDNA become necessary, relying on sequencing of the hypervariable region itself is not very practical in a forensic context since it is very time consuming, labour intensive and relatively expensive. Due to this scenario, development of a rapid and robust mtDNA analysis method become very crucial.

Minisequencing has been seen as a promising method for mtDNA analysis. Multiplex solid-phase fluorescent minisequencing is one of the validated minisequencing methods for forensic casework (Tully *et al.*, 1996; Morley *et al.*, 1999). The method is based on single nucleotide extension of the minisequencing primer with a labelled ddNTP (Figure 1.2). The fluorescent labels were then detected in an automated 4-colour sequencer, resulting a peak like profile. This method had so far optimised to identify ten substitution polymorphisms and two length polymorphisms of the non-coding region simultaneously (Morley *et al.*, 1999). Probability of random match calculated using this method from databases of 152 British Caucasians and 103 British Afro-Caribbeans was 0.054 and 0.026, respectively (Tully *et al.*, 1996). The random match probability calculated using this method was however higher than using sequence analysis of HVI and HVII region for other European Caucasians (Budowle *et al.*, 1999b; Baasner *et al.*, 1998). More polymorphic sites should be examined in order to improve the random match probability in minisequencing analysis.

Figure 1.2: Schematic diagram of the multiplex solid-phase minisequencing (Morley *et al.*, 1999).

1. 

Three fragments of the mtDNA non-coding region were amplified by triplex PCR using biotinylated primers (only one of the primers is biotinylated).

2. 

One of the amplified strands is immobilised on Dynabeads. The remaining PCR products, along with dNTPs were wash away. Each minisequencing primer (shown in red) was designed to be one nucleotide short at the 3′ end, immediately before the polymorphic site to be detected. Polymerase enzyme extended the primer with a single fluorescently labelled ddNTPs.

3. ATCCGAATTCCCGATGGGTTA
TAACGTCGAAC
} Labelled minisequencing primers

The extended and labelled minisequencing primers were eluted from immobilised template and loaded to an automated 4-colour sequencer (ABI 377, ABI 373 or ABI 310) for the fluorescent labels detection.

1.4 Application of mtDNA in Population Studies

Important genetic records of the ancient populations history lies within the mitochondrial genome. The study of mtDNA hit popular consciousness in 1987 by the debate on modern human origins and the African 'mitochondrial Eve' (Cann *et al.*, 1987). Special features of maternal inheritance and rapid rate of mutations had made mtDNA as a valuable tool in the study of evolutionary genetics of modern humans and also to understand the demographic history of human populations (Wilson *et al.*, 1985). The mutations that have struck throughout human history trace the maternal genealogy and through this genealogy we can attempt to deduce something about prehistoric process (Macaulay *et al.*, 1999).

Sequence variations of mtDNA have provided a magnified view of genetic diversity between individuals within populations and also between different populations (Cann *et al.*, 1987). Mitochondrial genome is the locus where most information on DNA sequence diversity in humans is currently available (Paabo, 1999). Previous analyses have also shown that most human mtDNA sequence variation had accumulated sequentially along radiating maternal lineages from sets of mtDNA founders, during and after the process of human colonisation of different geographical regions of the world (Torroni *et al.*, 1998). This is the reason why mtDNA haplogroups defined by high-resolution RFLP analysis are often found to be geographically and ethnically specific (Wallace, 1995) (Table 1.2).

Studies of mtDNA polymorphisms in human populations were based on restriction enzyme analysis (Cann *et al.*, 1987) and sequencing of the hypervariable regions

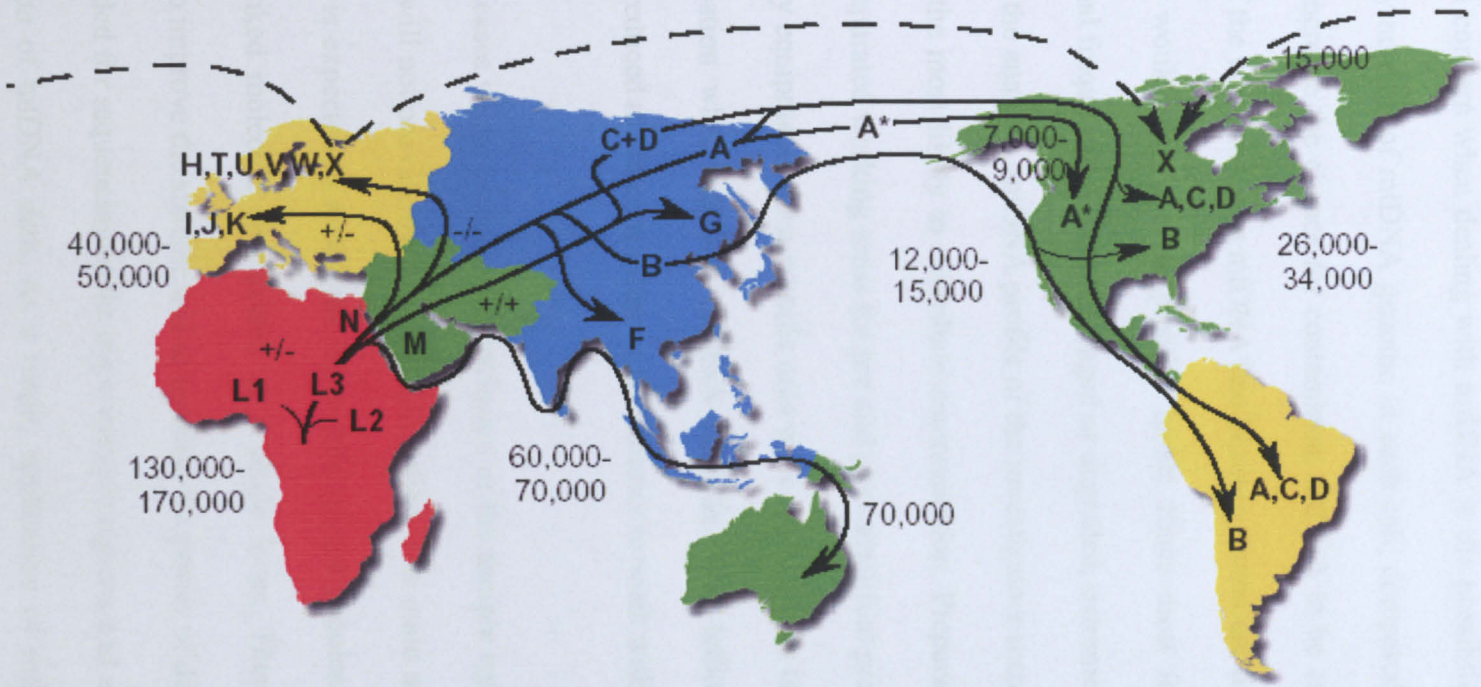
(Vigilant *et al.*, 1991). Most comprehensive information was gained from the RFLP analysis of the coding region (Torroni *et al.*, 1998). The earliest work involved the digestion of the entire molecule either using a single restriction enzyme in large sample number (Denaro *et al.*, 1981) or more enzymes in less samples (Brown, 1980). Until today, several haplogroups have been assigned for specific populations. Almost 95% of the Western Eurasian mtDNA pool, including Mediterranean Africa was comprised of haplogroup H, I, J, K, T, U, V, W and X (Torroni *et al.*, 1996). Haplogroups M, B, F and A were found to be more familiar within the Mongoloid populations while the Sub-saharan Africa belong largely to haplogroup L (Watson *et al.*, 1997). The world mtDNA human migration patterns and each continent specific haplogroups are shown in Figure 1.3.

Table 1.2: Common Continent Specific mtDNA Variants.

(As edited from www.mitomap.org)

Continent	Haplogroup	Defining Polymorphisms
Africa	L	3592 +*Hpa* I
Europe	H	7025 -*Alu* I
	I	1715 -*Dde* I; 10028 +*Alu* I; 4529 -*Hae* III; 8249 +*Ava* II; 16389 +*Bam*H I/*Mbo* I
	J	13704 -*Bst*N I; 16065 -*Hin*f I
	K	9052 -*Hae* II/*Hha* I
Asia	A	663 +*Hae* III
	B	8271-8291 9bp deletion; 16517 +*Hae* III
	F	12406 -*Hpa* I/*Hin*c II; 16517 +*Hae* III
	M	10394 +*Dde* I; 10397 +*Alu* I
	M-C	13259/13262 +*Hin*c II/-*Alu* I
	M-D	5176 -*Alu* I
America	A (Amerinds)	663 +*Hae* III
	A (Na-Dene)	663 +*Hae* III; 16329 -*Rsa* I
	B (Amerinds)	8271-8291 9bp deletion; 16517 +*Hae* III
	M-C (Amerinds)	13259/13262 +*Hin*c II/-*Alu* I
	M-D (Amerinds)	5176 -*Alu* I

Figure 1.3: The world mtDNA human migration patterns and each continent specific haplogroups (www.mitomap.org).



+/+, +/- or -/- = 10304/10397 *Dde* I/*Alu* I
* = -*Rsa* I 16329

Mutation rate: 2.2-2.9% / MYR
Time estimated are YBP

1.5 Issues in mtDNA Typing

The main concern when dealing with mtDNA is the possibility of contamination. The high copy number of mtDNA genome in each cell, compared to nuclear genome, will actually increase the chance for contaminant mtDNA to be amplified. In severe cases, profile of the contaminating mtDNA had become prominent from the actual sample and therefore would yield a false positive result. Since most forensic samples are very limited and frequently severely damaged or degraded, extreme care must be taken when handling the analysis. mtDNA profile of the investigators seems to be compulsory since they are the most likely to contribute contamination. Preparation of negative controls, blanks, separated working areas for pre and post-amplified products and designations of laboratory equipments for a specific area were among steps taken to monitor and avoid contamination when handling mtDNA analysis. This tedious and extra works had actually reduced the interest of many laboratories to work with mtDNA.

In crime cases where exclusion or inclusion of the sample typed was the main concern, mtDNA will not provide definitive identification. The main argument is that the profile obtained is expected to match the maternally related peoples. In addition, mtDNA is a single linked molecule and treated as a single locus. Therefore a large database is needed to improve the statistic and to increase the power of discrimination. Expense and time needed for sequencing of the hypervariable regions had actually delay an enormous production of mtDNA data. As a result, application of mtDNA in forensic cases is limited by lacking of reasonable number of information in the databases. To combat this problem, analysis of single nucleotide polymorphisms (SNPs) of the coding region had been seen as a useful complement to the information gained from the hypervariable

regions. As new rapid, large scale and high throughput technologies such as pyrosequencing, microarrays and minisequencing being introduced for the interrogation of SNPs, this might increase the utility of mtDNA in forensic.

1.6 Aims Of The Project

The aim of this study is to examine the mtDNA of the Modern Malay and Orang Asli populations in Peninsular Malaysia. Hitherto, very limited information of mtDNA data from Malaysian populations is available. The underrepresented mtDNA data and lack of relevant databases had also setback the application of mtDNA in most forensic caseworks in Malaysia. Examination of mtDNA variations and classification into specific mtDNA haplogroups will help in determining the population history and will also be useful in forensic applications. Comparison of mtDNA variation pattern between the Modern Malay and Orang Asli, which is the largest population and the aborigine population of Peninsular Malaysia, respectively, will also help to resolve the relationship of these two populations since both are believed to be among the earliest people to inhibit Peninsular Malaysia. In order to achieve reliable information in the most efficient way, several methods of mtDNA analysis will be used. Both coding and non-coding region polymorphisms will be employed in the analysis using RFLP and direct sequencing method, respectively. In addition to these conventional methods, a new rapid and robust technique known as minisequencing will be tested and optimised for the interrogation of the coding region SNPs. As an addition, the autosomal STR analysis will be perform for both Modern Malay and Orang Asli samples. This will enhance the number of data available for the STR database of the Peninsular Malaysia populations.

1.7 Milestones of the project

1) Samples collection from unrelated individuals of the Modern Malay and Orang Asli population in Peninsular Malaysia.

2) DNA extraction and quantification

3) Direct sequencing of the HVI region: Amplification of HVI region

Automated fluorescent sequencing

Reporting polymorphisms

4) RFLP analysis of the coding region: Primer design

Amplification of the target region

Restriction digest

Agarose gel electrophoresis

Data interpretation

5) Minisequencing: Primer design (minisequencing primers)

Optimisation of the PCR condition

Amplification and purification of minisequencing templates

SNaPshot singleplex reaction and optimisation

SNaPshot multiplex reaction and optimisation

6) Analysis of nuclear DNA polymorphisms

CHAPTER 2: MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Chemicals

All chemicals used in the experiment were obtained from Sigma Chemical Co. USA, Bio-Rad Laboratories Ltd. UK and Fisons Scientific Equipment, UK.

2.1.2 Enzymes

2.1.2.1 Restriction Enzymes

All restriction enzymes used were purchased from Promega Corp., USA. Each enzyme was provided with 10X reaction buffer and 25 mM Magnesium chloride solution.

2.1.2.2 Shrimp Alkaline Phosphatase (SAP)

This enzyme was used to dephosphorylates all remaining dNTPs in the post PCR reaction. It was purchased in 5U/μl concentration from USB Corporation, USA

2.1.2.3 *Exonuclease* I (*Exo* I)

When necessary, *Exo* I was used to degrade all primers after DNA amplification was completed. It was purchased from USB Corporation, USA at concentration of 10U/μl.

2.1.2.4 Ampli*Taq*® Gold Enzyme

This enzyme was purchased from Applied Biosystems, USA. Each pack was enclosed with a tube of 5U/µl of Ampli*Taq*® Gold Enzyme, 10X reaction buffer and 25 mM Magnesium chloride solution.

2.1.3 ReddyMix™ PCR Master Mix

ReddyMix™ PCR Master Mix was purchased from ABgene, UK. This ready-to-use master mix was used in all PCR reactions. It was designed to contain all components required for PCR reactions, including the dye and precipitant to facilitate gel loading. Each vials contains 1.8 ml of 1.1X working concentration PCR Master Mix which sufficient for 40 x 50 µl reactions. The final reaction volume of 50 µl, with primers and template added contained the following components: 1.25 units *Taq* DNA Polymerase, 75 mM Tris-HCL (pH 8.8 at 25 °C), 20 mM $(NH_4)_2SO_4$, 1.5 mM $MgCl_2$, 0.01% Tween®20, and 0.2 mM each of dATP, dCTP, dGTP and dTTP.

2.1.4 Collection Swabs

Sterile collection swabs, which used to collect buccal cells from donors were manufactured by I.A.S.A, Spain.

2.1.5 Kits

A number of kits were used in this study. These kits were purchased from several manufacturers and used according to the methods described by the manufacturer.

2.1.5.1 Puregene® DNA Extraction Kit

This kit was purchased from Gentra Systems (Flowgen, Novara Group Limited, UK) and being used for DNA extraction from buccal swabs. Each kit consisted of RBC Lysis Solution, Cell Lysis Solution, Protein Precipitation Solution and DNA Hydration Solution.

2.1.5.2 Wizard® *Plus* Minipreps DNA Purification System

The use of this system was according to the protocol provided by the manufacturer, Promega Corporation, USA. Each system contained the following reagents, which sufficient for 50 standard minipreps from 1-3 ml culture: 10 ml Cell Resuspension Solution, 10 ml Lysis Solution, 20 ml Neutralisation Solution, 50 ml Wizard® Minipreps DNA Purification Resin, 50 ml Column Wash Solution, 50 Wizard® Minicolums and 50 Syringe Barrels (3 ml).

2.1.5.3 pGEM-T® Easy Vector Systems

Cloning of PCR fragments were done using this kit, which was also purchased from Promega Corporation, USA. Each kit includes: 1.2 µg pGEM-T® Easy Vector (50 ng/µl), 12 µl Control Insert DNA (4 ng/µl), 100U T4 DNA Ligase, 200 µl 2X Rapid Ligation Buffer for T4 DNA Ligase and 1.2 ml JM109 Competent Cells, High Efficiency (6 x 200 µl).

2.1.5.4 ABI PRISM® SNaPshot™ Multiplex Kit

This kit was purchased from Applied Biosystems, USA and contained the following items: SNaPshot™ Multiplex Ready Reaction Mix (Amplitaq® DNA Polymerase, Fluorescently labelled ddNTPs and reaction buffer), SNaPshot™ Multiplex Control Primer mix (20A primer 0.05 pmol/µL, 28G/A primer 0.10 pmol/µL, 36G primer 0.05 pmol/µL, 44T primer 0.30 pmol/µL, 52C/T primer 0.30 pmol/µL and 60C primer 0.30 pmol/µL) and SNaPshot™ Multiplex Control Template (Amplicon from CEPH DNA).

2.1.5.5 QuantiBlot® Human DNA Quantitation Kit

The above kit was purchased from Applied Biosystem, USA. Each kit consisted of: QuantiBlot® D17Z1probe (220 µl), HRP-SA Enzyme conjugate (2.0 ml), Bromothymol Blue solution (200 µl), QuantiBlot® DNA standard A (2 ng/µl), QuantiBlot® DNA Calibtrator 1 (0.7 ng/µl) and QuantiBlot® DNA Calibrator 2 (0.1 ng/µl).

2.1.5.6 STR PowerPlex® 16 System

This kit consists of pre-amplification and post-amplification components, which stored and used in a totally separated area to avoid cross contamination of the components. The pre-amplification box (with yellow label) contained 1 x 300 µl Gold ST*R 10X Buffer, 1 x 250 µl PowerPlex® 16 10X Primer Pair Mix and 25 µl 9947A DNA (10 ng/µl). The post-amplification box (with magenta label) consist of 1 x 25 µl PowerPlex® 16 Allelic Ladder Mix, 1 x 150 µl Internal Lane Standard (ILS) 600 and 1 x 1 ml Blue Dextran Loading Solution.

2.1.6 PCR Primers

PCR primers were synthesised by Invitrogen Life Sciences, UK. The primers were designed manually by using The Cambridge Reference Sequence (Andrews *et. al.*, 1999). All PCR primers were listed in Table 2.1, corresponded to the region amplified. Each primer was diluted to a final concentration of 100 pm/µl and kept at -20 °C for longer storage.

Table 2.1: The sequence of all mitochondrial primers used in PCR and the corresponding region/polymorphism amplified.

Table 2.1.1: Sequence of each primer used in sequencing analysis.

Primers	Sequence (5′ to 3′)	Region
L15996	CTC CAC CAT TAG CAC CCA AAG C	HV1 region
H16431	CGA GGA GAG TAG CAC TCT TG	
L15926	TAC ACC AGT CTT GTA AAC C	HV1 region

Table 2.1.2: Sequence of each primer used in RFLP analysis and amplification of minisequencing templates.

Primers	Sequence (5′ to 3′)	RFLP site
9911-F	CGA AGC CGC CGC CTG ATA CTG G	+10394/+10397 *Dde* I/*Alu* I
11873-R	TGG GGG GTA AGG CGA GGT TAG C	
10291-F	TCC TTT TAC CCC TAC CAT GAT	+10394/+10397 *Dde* I/*Alu* I
10556-R	ATT ATT CCT TCT AGG CAT AGT AG	
4887-F	TGA CAA AAA CTA GCC CCC ATC T	-5176 *Alu* I
5160-F	GCA CCA CGA CCC TAC TAC TAT	
5442-R	GCG ATG AGT GTG GGG AGG AA	
12978-F	CAA GCC TCA CCC CAC TAC TA	+13262 *Alu* I
13773-R	GGT AGA GGG GGA TTG TTG TT	
12233-F	GCT CAC AAG AAC TGC TAA CTC	-12406 *Hinc* II
12630-R	ACA GTG AGA ATT CTA TGA TGG AC	

Primers	Sequence (5′ to 3′)	RFLP site
1643-F	AAC ACA AAG CAC CCA ACT TAC ACT TAG GA	-1715 *Dde* I
1874-R	CTT TGG CTC TCC TTG CAA AGT	
6889-F	GAA GCA ATA TGA AAT GAT CTG C	-7025 *Alu* I
7115-R	GGC GTA GGT TTG GTC TAG G	
8215-F	Fam-ACA GTT TCA TGC CCA TCG TC	9-bp deletion
8385-R	GTA ATT ATG GTG GGC CAT ACG G	
396-F	CCA GCC TAA CCA GAT TTC AA	+663 *Hae* III
1067-R	GGG GTA TCT AAT CCC AGT TT	
6889-F	GAA GCA ATA TGA AAT GAT CTG C	-7598 *Hha* I
8385-R	GTA ATT ATG GTG GGC CAT ACG G	
8845-F	CCT AGC CAT GGC CAT CC	-8994 *Hae* III
9163-R	TTA CTA GAA GTG TGA AAA C	
14394-F	CTC CAT CGC TAA CCC CAC	+14465 *Acc* I
14597-R	CTT CTA AGC CTT CTC CTA T	
12104-F	CTC AAC CCC GAC ATC ATT ACC	+12308 *Hinf* I
12309-R	TTA CTT TTA TTT GGA GTT GCA CCA AGA TT	
14560-F	GAC CAC ACC GCT AAC AAT CAA TAC	-14766 *Mse* I
15081-R	GTT TCT GAG TAG AGA AAT GAT CCG	
4491-F	GTC ATC TAC TCT ACC ATC TT	-4577 *Nla* III
4676-R	GAT TAT GGA TGC GGT TGC TTG	
8334-F	GAT TAA GAG AAC CAA CAC CTC	-8616 *Mbo* I
8880-R	GCG AAA GCC TAT AAT CAC TGT G	

Primers	Sequence (5′ to 3′)	RFLP site
2107-F	GGA ACA GCT CTT TGG ACA CTA G	+2349 *Mbo* I
2637-R	GCC ATT CAT ACA GGT CCC TAT	
13173-F	CTT AGG CGC TAT CAC CAC TCT G	+13367 *Ava* II/ +13367 *Mbo* I
13580-R	GCG ATG AGA GTA ATA GAT AGG GC	
11543-F	CTT GTA CTA TCC CTA TGA GGC	+11718 *Hae* III
11910-R	GTT ACT AGC ACA GAG AGT TCT C	

2.2 METHODS

2.2.1 Sterilisation

All plastic ware, glassware, medium, deionised water and instruments that needed to be in sterile condition were autoclaved at 20 psi for 15 min. Pippete tips were sterilised in microwave oven for 30 min while the microfuge tubes were sterilised by exposing the inside wall of the tubes to UV light for 20 min.

2.2.2 Separate Working Areas

All PCR reactions and primer dilutions were carried out in a flow hood to avoid extraneous contamination. The flow hood was cleaned with decon prior and after all works. Post-PCR products were handled in a separate room to prevent contamination of the amplified DNA in the laboratory. A set of pipettes, laboratory coats, gloves and other equipment and consumables were designated to each working area so that they were not carried from one place to another which may caused cross over contamination.

2.2.3 DNA Extraction from Buccal Swabs

The buccal swabs were cut into sterile 1.5 ml screw cap tubes and labelled accordingly. A total of 400 µl of cell lysis solution was added to the swabs, vortexed for 1 min and then 20 µl proteinase K (20 mg/ml) was added. The mix was incubated at 60 °C for 2 h. At the end of incubation, 200 µl of the solution was transferred into new sterile 1.5 ml microfuge tubes and the other remaining solutions with the buccal swabs in it were kept

at -20 °C. To each solution in the microfuge tubes, 100 µl of protein precipitation solution was added and vortexed thoroughly for about 1 min. The microfuge tubes were centrifuged at 12,000*g* for 10 min and the supernatant was transferred into another microfuge tubes containing 300 µl isopropanol. The tubes were inverted 4 times to mix the solution and centrifuged at 12,000*g* for another 10 min. Supernatant was discarded carefully and 100 µl of 80% ethanol was added into the tubes to wash the pellet. The tubes were air dried for about 15 min before 50 µl of DNA hydrating solution was added. Tubes were incubated at 37 °C for one hour to make sure all DNA had dissolved and were then stored at -20 °C.

2.2.4 DNA Quantitation

Determination of DNA concentration of the buccal swab extractions was done using the QuantiBlot® Human DNA Quantitation Kit. This kit provided rapid quantitation of human DNA. The entire protocol can be completed within 2 hours and sensitive enough to quantitate from 0.15 to 10 ng of human DNA. This particular kit allows two detection methods, either colorimetric or chemiluminescent detection. The quantity of each DNA samples in either colorimetric or chemiluminescent detection was determined by comparison to the signal intensity of human DNA standard (2 ng/µl), which is provided with the kit. The colorimetric detection was chosen since a dark room equipped with X-ray film development was not available.

A series of dilutions were carried out in order to have a DNA standard with a concentration range between 2 ng/µl to 0.03125 ng/µl (Table 2.2). Each set of DNA samples to be quantitated must be accompanied by these DNA standards, along with

DNA Calibrator 1 and DNA Calibrator 2. To prepare the samples for quantitation, 2 µl of each DNA sample was added into a labelled 1.5 ml microfuge tube containing 150 µl spotting solution (0.4 M NaOH, 25 mM EDTA and 0.00008% Bromothymol Blue). For DNA standards and DNA calibrators, 5 µl from each dilution was added to 150 µl spotting solution.

Two pieces of Biodyne® B membrane were cut with the size of 11.0 cm x 7.9 cm each. A small notch was cut in the upper right corner of each membrane to mark the orientation. The membranes were placed in hybridisation trays containing 50 ml pre-wetting solution (0.4 M NaOH, 25 mM EDTA) and incubated at room temperature between 1 to 30 min. After the incubation period, each membrane was transferred using clean forceps onto the gasket of the slot blot apparatus. The top plate of the slot blot apparatus was placed on top of the membrane. The vacuum source was then turned on with the sample vacuum on the slot blot apparatus being turned off and the clamp vacuum turned on. The top plate was pushed down to ensure a formation of tight vacuum seal.

Each DNA sample in the spotting solution was then dispensed slowly into a different well of the slot blot apparatus. After all the samples were transferred, the sample vacuum was slowly turned on. The vacuum was left on until all the samples were drawn through the membrane and a blue band was formed. The membranes were then removed from the slot blot apparatus and immediately soaked in 100 ml of pre-warmed hybridisation solution (5X SSPE, 0.5% w/v SDS). A total of 5 ml of H_2O_2 was added and the membranes were incubated (pre-hybridisation treatment) at 50 °C in a rotating water bath (50 to 60 rpm) for 15 min. After the incubation period, the solution was

poured off and replaced with 30 ml pre-warmed hybridisation solution. With the tray tilted at one side, 20 µl of QuantiBlot D17Z1 probe was added to the hybridisation solution. The membranes were incubated again at 50 °C in a rotating water bath for another 20 min.

The membranes were then rinsed briefly in 100 ml pre-warmed wash solution (1.5X SSPE, 0.5% w/v SDS). The solution was poured off and 30 ml of pre-warmed wash solution was added. With the tray tilted at one side, 180 µl of Enzyme Conjugate HRP-SA was added to the hybridisation solution. The membranes were further incubated at 50 °C in a rotating water bath for 10 min. After the incubation ended, the membranes were rinsed thoroughly for 1 min with 100 ml of pre-warmed wash solution by rocking on an orbital shaker (100 to 125 rpm). This procedure was repeated three times before the membranes were rinse briefly in 100 ml Citrate Buffer (0.1 M Sodium Citrate, pH 5.0).

The Colour Development Solution was prepared by adding 1.5 ml Chromogen:TMB solution (from Applied Biosystems) and 30 µl of 3% H_2O_2 into 30 ml Citrate Buffer. This solution was added immediately to the membranes and shaken (50-60 rpm) at room temperature for 20-30 min. The colour development was stopped by washing the membranes in 100 ml deionised water for three times. The membranes were covered with Saran Wrap® and photographed instantly to prevent losing colour intensity.

Table 2.2: Serial dilution of DNA standard in TE buffer.

Tube	Total from stock (µl)	TE buffer (µl)	Total transferred to the next tube (µl)	Concentration (ng/µl)
A	120	-	60	2
B	-	60	60	1
C	-	60	60	0.5
D	-	60	60	0.25
E	-	60	60	0.125
F	-	60	60	0.0625
G	-	60	60	0.03125

*After each transfer, the tubes with new dilution mix were vortexed to blend the contents thoroughly.

2.2.5 Sequencing Analysis of HVI Region

2.2.5.1 Polymerase Chain Reaction (PCR)

The HVI region was amplified using a pair of primers, L15996 and H16431. PCR reaction was done in 0.2 ml PCR tubes with the following mixtures: 1 µl of each forward and reverse primer (10 pm/µl), 2 µl of DNA template (5ng/µl) and 46 µl of PCR ReddyRun™ Master Mix. The following PCR condition was used: 94 °C for 5 min, 30 cycles of 94 °C for 30 s, 54 °C for 30 s, 72 °C for 45 s and final extension step at 72 °C for 7 min. PCR products were kept at 4 °C in a separate fridge from the pre-PCR components to avoid contamination.

2.2.5.2 Agarose Gel Electrophoresis

Agarose gel (1.2%) was prepared by adding 0.6 g Agarose LE (Flowgen Ltd, UK) into 50 ml 1X TBE buffer. The mixture was boiled in a microwave oven until all gel was dissolved. The solution was allowed to cool to 60 °C before adding 1 µl ethidium bromide (10 mg/ml). After the gel was set in the casting tray, 2 µl of the PCR products and 1 µl of 1 kb DNA ladder (1 µg/µl) were loaded into different wells and electrophoresed in 1X TBE buffer at 80 V for 45 min. The gel was photograph and used to estimate the concentration of the PCR products by comparing the intensity of each band to the 1.6 kb band of the 1 kb DNA ladder. Since the PCR products were used as templates in the sequencing reactions, estimation of the concentration for each product was necessary.

2.2.5.3 Automated Sequencing

Purification and sequencing reaction of the PCR product was performed by the Molecular Biology Science Unit, Institute of Molecular Biology, University of Glasgow. The primers used for the sequencing reaction were listed in table 2.1.1. PCR fragment was sequenced using ABI Prism Big Dye® Terminator Cycle Sequencing Kit on an ABI 377 sequencer. Each sequence obtained was then aligned with the Cambridge Reference Sequence (Andrews *et al.*, 1999), starting from position 16001 to 16400 using GeneJockey Sequence Processor software (Taylor, 1990) in order to detect the polymorphisms.

2.2.6 Cloning of PCR Fragments

PCR fragments were cloned using the pGEM®-T Easy Vector System. PCR products were ligated to the pGEM®-T Easy Vector in the following reaction: 5 µl of 2X Rapid Ligation Buffer, 1 µl of pGEM®-T Easy Vector, 1 µl of PCR product, 1 µl of T4 DNA Ligase (3 Weiss units/µl) and 2 µl of deionised water. The reaction was mixed by pipetting the contents for few times and incubated at 4 °C overnight.

LB plates with ampicillin (50 µg/ml), X-gal (40 µg/ml) and IPTG (40 µg/ml) were prepared prior to the transformation. The JM109 tube was thawed on ice and the content was mixed by flicking the tube gently. 50 µl of the competent cells were then added to 2 µl of the ligation mix. The tubes were gently flicked again to mix the solution and were placed on ice for 20 min. The cells were then heat shocked at 42 °C for 50 s and immediately returned to ice for another 2 min. A total of 950 µl LB medium was added

to the tubes containing cells transformed with the ligation reaction and further incubated at 37 °C for 1.5 h with shaking (approximately 150 rpm). At the end of the incubation period, the cells were pelleted by centrifugation at 1000*g* for 10 min. The cells were resuspended in 200 µl of LB medium and plated on 2 LB plates (100 µl each). The plates were incubated overnight at 37 °C in inverted position.

2.2.7 Screening and Purification of Cloned PCR Fragments

Transformants from the above method (2.2.6) were screened for inserts by selecting the white colonies formed on the LB plates after overnight incubation. The colonies were picked using sterile toothpicks and transferred onto grid LB/Amp plates. The remaining cells were inoculated into 5 ml LB broth containing 50 µg/ml ampicillin and incubated at 37 °C overnight with shaking (200 rpm). A total of 1.5 ml from the incubated products were transferred into sterile microfuge tubes and centrifuged at 1000*g* for 30 s to pellet the cells. The supernatant was discarded and the tubes were blotted up side down on paper towel to remove access media.

Purification of the cloned plasmids was performed using the Wizard® Plus Minipreps DNA Purification Systems. The pelleted cells were resuspended in 200 µl of Cell Resuspension Solution. After the cells were completely dissolved, 200 µl of Cell Lysis Solution was added and mixed by inverting the tubes 4 times. A total of 200 µl of Neutralising Solution was added and mixed again by inverting the tubes few times until a white clump was formed. The lysate was centrifuged at 1000*g* for 5 min to precipitate the white clump and the supernatant was transferred into a syringe barrel containing 1 ml of purification resin. The syringe plunger was carefully inserted and the slurry was

gently pushed through the Minicolumn, which was earlier attached to the syringe barrel. The Minicolumn was rinsed with 2 ml Column Wash Solution before being transferred to a 1.5 ml microcentrifuge tube and centrifuged at 1000*g* for 2 min. The Minicolumn was then transferred to a new 1.5 ml microcentrifuge tube and added with 50 µl nuclease-free water. After 1 min incubation, the Minicolum was centrifuged again at 1000*g* for 20 s to elute the DNA. The microcentrifuge tube containing DNA was stored at -20 °C.

2.2.8 RFLP Analysis of mtDNA Coding Region

2.2.8.1 Primer Design

All primers used to amplify the target region were design manually by using the Cambridge Reference Sequence (listed in Table 2.1.2). Mutation sites selected for the RFLP analysis were obtained from established publications (see chapter 5) and Mitomap site (www.mitomap.org). The selected oligo sequences were between 18 to 30 bp long with at least 40% of GC content. The forward and reverse primers were designed carefully in order to avoid primer dimer and hairpin structure formation, which can interrupt the PCR efficiency.

2.2.8.2 Polymerase Chain Reaction

PCR reaction was prepared in 0.2 ml thin-walled tube with the following mixture: 1 µl of each primer (10 pmoles/µl), 2 µl of DNA template (5 ng/µl) and 11 µl of PCR ReddyRun™ Master Mix. PCR condition used was the same as described in 2.2.5.1

except for different annealing temperature was used for different set of primers, as listed in Table 2.3. In order to check the quality and quantity of the amplification product, 2 µl of the PCR product was electrophoresed in 1.2% agarose gel at 80 V for 45 min. The remaining product was stored at 4 °C.

2.2.8.3 Restriction Digest

The following restriction enzymes were used for the digestion: *Alu* I, *Dde* I and *Hinc* II. A master mix was prepared for each set of digestion, which contained 2.5U of enzyme, 1X digestion buffer and topped up to the required volume with deionised water (Table 2.4). A total of 2 µl from the master mix was added to 13 µl of the PCR product to make a final reaction volume of 15 µl. The reaction mix was incubated at 37 °C for 5 h. At the end of incubation, 5 µl of the reaction mix was electrophoresed in 1.2% agarose gel at 80 V for 45 min. If all the expected bands were visualised, the gel was photographed and kept for further reference.

Table 2.3: Annealing temperature used in the amplification of the target region for RFLP analysis.

Primer	Annealing Temperature (°C)
1643-F 1874-R	52
3409-F 3700-R	46
4887-F 5160-F 5442-R	56
6889-F 7115-R	56
8215-F 8385-R	60
9911-F 11873-R	54
9932-F 10088-R	52
10291-F 10556-R	48
12233-F 12630-R	50
12978-F 13773-R	56

Table 2.4: Master Mix for Restriction Digestion Analysis.

Component	Volume per reaction (µl)
Restriction Enzyme	0.25
1X Reaction Buffer	1.5
Deionised water	0.25

2.2.9 STR PowerPlex® 16 Analysis

The PowerPlex® 16 System used in the STR Analysis provides all materials (except for AmpliTaq Gold® DNA polymerase) necessary for the amplification of STR regions of purified genomic DNA. This system was specially design for use with ABI PRISM® 310 Genetic Analyser and also compatible with the ABI PRISM® 3100 Genetic Analyser and 377 DNA Sequencer. All Modern Malay and Orang Asli samples were analysed using this system.

2.2.9.1 STRs Amplification

A master mix was prepared for the amplification reactions as listed in Table 2.5. A total of 10.5 µl from the master mix was added to 2 µl DNA template (1 ng/µl), making a final reaction volume of 12.5 µl. For the positive and negative amplification control, 1 ng of 9947A DNA and 2 µl Nuclease-Free Water were added to 10.5 µl PCR master mix, respectively. The amplification process took place in a Perkin-Elmer Thermal Cycler Model 2700 using the cycling condition stated in Table 2.6.

2.2.9.2 GeneScan Analysis

The amplified products were analysed using ABI PRISM® 310 Genetic Analyser. For each injection, 0.5 µl Internal Lane Standard 600 (ILS 600) and 9 µl deionised formamide were added to 1 µl amplified products. The same amount of ILS 600 and formamide were added to 1 µl PowerPlex® 16 Allelic Ladder Mix. The samples and allelic ladder were denatured by heating at 95 °C for 3 min and snapped cold on ice for

another 3 min. The tubes were then assembled in the autosampler tray and placed in the instrument. The capillary electrophoresis was performed using the GS STR POP4 (1 ml) A module and the results were viewed using the GeneScan® Analysis Software.

2.2.9.3 PowerTyper™ 16 Macro

The assignment of genotypes for the Modern Malay and Orang Asli samples were carried by the PowerTyper™ 16 Macro. This programme was created to facilitate analysis of the data generated by the PowerPlex® 16 System. To perform analysis, the GeneScan® project files were imported into this software. The Internal Lane Standard was first checked to confirm that the fragment size was correct. The 'POWER' macro will then identify each allele in the ladder sample and calculates offsets for all of the loci.

Table 2.5: Master Mix for STR PowerPlex® Kit Reaction.

Component	Volume (µl) per reaction
Primer pair mix	1.25
STR Gold Buffer	1.25
Ampli*Taq*® Gold Enzyme	0.4
Deionised water	7.6
Total volume	10.5

Table 2.6: Thermal Cycling Parameters for STR PowerPlex® 16 Amplification.

Step	Temperature (°C)	Time (min)	No of Cycles
Initial Incubation	95	11	1
Denaturation	96	2	1
Amplification 1	94 60 70	1 1 1.5	10
Amplification 2	90 60 70	1 1 1.5	20
Final Extension	60	30	1
Soak	4	-	-

2.2.10 Minisequencing

Minisequencing was performed to interrogate the single nucleotide polymorphisms (SNPs) responsible for the existence or absence of restriction sites within the coding region of mtDNA. Detection of SNPs was carried out using the ABI PRISM® SNaPshot™ Multiplex Kit. The chemistry of this kit was based on the dideoxy single-base extension of unlabeled oligonucleotide primers.

2.2.10.1 Minisequencing Templates

Before the single base extension was possible, PCR was performed using primers listed in Table 2.7 to generate templates. Each PCR reaction contained 1 µl of each forward and reverse primer (10 pm/µl), 2 µl of DNA template (5 ng/µl) and 13 µl of ReddyMix™ PCR master mix to make a final volume of 17 µl. PCR conditions used are described in section 2.2.5.1 and the annealing temperature used are listed in Table 2.7.

2.2.10.2 Post PCR Treatment

The post-PCR products also include excess PCR primers, dNTPs, enzyme and buffer component. To avoid participation of primers and dNTPs in subsequent primer-extension reactions, these materials were removed by digestion with *Exo* I and SAP. A total of 5 µl of SAP (1 unit/µl) and 0.2 µl (10 unit/µl) of *Exo* I was added to the post PCR reaction. Each enzyme was added one at a time and mixed thoroughly each time. The mixture was incubated at 37 °C for one hour and incubated for another 15 min at 75 °C to inactivate the enzymes.

Table 2.7: PCR primers for amplification of the minisequencing templates, annealing temperatures used, the target polymorphic sites, nucleotide changes and RFLP sites created by the polymorphisms.

Primers	Annealing temperature (°C)	Polymorphic site	Nucleotide changes	RFLP site
396F 1067R	50	663	A to G	+663 *Hae* III
2107F 2637R	56	2352	T to C	+2349 *Mbo* I
4491F 4676R	56	4580	G to A	-4577 *Nla* III
L5160 H5442	56	5178	C to A	-5176 *Alu* I
L6889 H7115	56	7028	C	-7025 *Alu* I
L6889 H8385	56	7598	G to A	-7598 *Hha* I
8334F 8880R	56	8618	T to C	-8616 *Mbo* I
8845F 9163R	48	8994	G to A	-8994 *Hae* III
10291F 10556R	48	10398 10400	A to G C to T	+10394 *Dde* I +10397 *Alu* I
11543F 11910R	56	11719	G to C	+11718 *Hae* III
12104F 12309R	50	12308	A to G	+12308 *Hinf* I
L12233 H12630	50	12406	G to A	-12406 *Hinc* II
L12978 H13773	56	13263	A to G	+13262 *Alu* I
13173F 13580R	56	13368	G to A	+13367 *Ava* II +13367 *Mbo* I
14394F 14597R	48	14470	T to C	+14465 *Acc* I
14560F 15081R	56	14766	C	-14766 *Mse* I

2.2.10.3 Minisequencing Primers

A total of 17 minisequencing primers (Table 2.8) were design manually using the Cambridge Reference Sequence (Andrews *et al.*, 1999). These primers were synthesised and HPLC purified by Invitrogen Life Sciences, UK. Each minisequencing primer was designed to be just one nucleotide shorter from the polymorphic site and the length of the primers were differed by 2-4 nucleotides to avoid overlapping between the SNaPshot products on the electropherogram. The length of each primer was modified by the addition of poly dA tails at the 5′end. All primers were carefully checked for possibility to form hairpin structure and also the formation of dimer between each primer. 3 primers, RA663G, RT8618C and RC10400T were design from the reverse strand since primer extension method did not allow flexibility with the respect to the 3′end location.

2.2.10.4 Singleplexing

Each minisequencing primer was tested in a singleplex reaction before being multiplexed. This was done in order to determine the best concentration and condition for the primer extension process. Reaction mix for this singleplex reaction contained 0.5 μM of the minisequencing primer, 5 μl of the SNaPshot Multiplex Ready Reaction Mix, 0.5 μl of the PCR template and 4 μl deionised water. SNaPshot mixture was kept on ice before being transferred to the thermal cycler to avoid high background. Thermal cycling process was conducted in a Perkin-Elmer Thermal Cycler Model 2700 with the following condition: 25 cycles of 96 °C for 10 min, 58 °C for 5 s and 60 °C for 30 s.

Table 2.8: Minisequencing primers used in the SNaPshot singleplex and multiplex reactions.

Primers	Sequences (5′ to 3′)	Size (bp)	Haplogroup
RA663G	CTA AGA GCT AAT AGA AAG GC	36	A
T2352C	CCG CAT AAG CCT GCG TCA GAT	52	L3e
G4580A	GAG TAG GCC TAG AAA TAA ACA T	30	V
C5178A	CTA TCT CGC ACC TGA AAC AAG	56	D
7028C	CGA CAC GTA CTA CGT TGT AGC	40	H
G7598A	CCT ATA TAT CTT AAT GGC ACA TGC A	40	E
RT8618C	GAT GAG ATA TTT GGA GGT GGG G	56	L3d
G8994A	CTC ATT CAA CCA ATA GCC CT	54	W
A10398G	CTA CAA AAA GGA TTA GAC TGA	48	M
RC10400T	GTT TAA ACT ATA TAC CAA TTC	52	M
G11719C	CAT TCT CAT AAT CGC CCA CGG	44	preHV
A12308G	TCC ATT GGT CTT AGG CCC CAA	36	U
G12406A	CCC ATC CTT ACC ACC CTC	44	F
G13368A	CCA TAC TAT TTA TGT GCT CCG G	48	T
A13263G	GCC TTC TCC ACT TCA AGT CA	60	C
T14470C	TAG CCA TCG CTG TAG TAT A	58	X
14766C	TGA CCC CAA TAC GCA AAA	32	HV

*Primer starts with R alphabet was design from the reverse strand.

*Different numbers of deoxyadenine (dA) were added at the 5′ end of each primer in order to achieve the size required.

2.2.10.5 Post-Extension Treatment

This procedure was performed to dephosphorylate all remaining ddNTPs from the post-extension mixture. If left untreated, unincorporated ddNTPs will co-migrate with the fragment of interest and cause interference. The 5′ phosphoryl group of ddNTPs was removed by adding 1 Unit of Shrimp Alkaline Phosphatase (SAP) to the post extension product and incubated at 37 °C for 1 h. SAP was deactivated by incubation at 75 °C for 15 min. The treated mixture was stored at -20 °C prior to electrophoresis on ABI 310 system.

2.2.10.6 Multiplexing

Two sets of multiplex reaction were set up for the minisequencing analysis (Table 2.9). The purified minisequencing templates for each set were premixed by adding 2 µl of each product in a microfuge tube. Minisequencing primers for each set were pooled together by adding 1 µl of each primer from the 100 pm/µl stock to a microfuge tube and making the total volume to 10 µl by adding deionised water. Each multiplex reaction contained the following components: 5 µl of the SNaPshot Multiplex Ready Reaction Mix, 3 µl of the pooled minisequencing templates, 1 µl of pooled minisequencing primers and 1 µl of deionised water. Thermal cycling process was conducted in a Perkin-Elmer Thermal Cycler Model 2700 with the following condition: 25 cycles of 96 °C for 10 s, 58 °C for 5 s and 60 °C for 30 s. The reaction mix for control reaction was listed in Table 2.10 and the thermal cycling condition used was the same as above, except for using 50 °C of annealing temperature. After thermal cycling

process was completed, post-extension product was treated as described in section 2.2.10.5.

2.2.10.7 GeneScan Analysis

SNaPshot product was electrophoresed on the ABI PRISM® 310 Genetic Analyser. The injection time and collection time encoded for GS POP-4 (1 ml) E5 Module was shorten from 5 s to 3 s and 24 min to 18 min, respectively. Prior to the electrophoresis of the SNaPshot products, a Matrix Standard Set DS-02 (dR110, dRGG, dTAMRA, dROX, LIZ) was run along with the control sample to create a matrix file compatible with the SNaPshot Multiplex Kit System. A loading cocktail was prepared by adding 0.5 µl of the SNaPshot product with 0.5 µl GeneScan-120 LIZ size standard and 9 µl Hi-Di formamide. The mixture was denatured at 95 °C for 5 min and placed on ice for another 3 min or until ready for loading into the 310 Genetic Analyser. The GeneScan-120 LIZ size standard analysis parameter was also created for automatic sizing of the SNaPshot product.

Table 2.9: Minisequencing primers designated for each set of SNaPshot multiplex reaction and the corresponding mtDNA haplogroup assigned by the polymorphic site.

Set	Minisequencing primers	mtDNA haplogroup
Multiplex 1	T2352C	L3e
	G4580A	V
	7025C	H
	RT8618C	L3d
	G8994A	W
	G11719C	preHV
	A12308G	U
	G13368A	T
	T14470C	X
	14766C	HV
Multiplex 2	RA663G	A
	C5178A	D
	G7598A	E
	A10398G	M
	RC10400T	M
	G12406A	F
	A13263G	C

Table 2.10: SNaPshot positive and negative control reaction mix.

Item	Positive Control (μl)	Negative Control (μl)
SNaPshot Multiplex Ready Reaction Mix	5	5
SNaPshot Multiplex Control Template	2	0
SNaPshot Multiplex Control Primer Mix	1	1
Deionised water	2	4
Total	10	10

CHAPTER 3: SAMPLES COLLECTION AND DNA EXTRACTION

3.1 Introduction

The sequence analysis of human mitochondria DNA (mtDNA) had been accepted and widely used in forensic science. In cases where the biological specimen that is available is severely decomposed or has been exposed to extreme environmental insult, such as high temperatures, mtDNA is often the only locus that can be successfully characterised. Even though mtDNA analysis is not as discriminating as nuclear DNA testing, it is still a reliable and powerful tool for human identification, especially when dealing with samples which do not contain genomic DNA like the hair shafts (Grzybowski, 2000). Besides the application in forensic cases, mtDNA polymorphisms had also been studied extensively as a tool to understand the demographic history of human populations.

Forensic applications of mtDNA mainly concern the examination of the HVI and HVII regions, which is highly variable between individuals. Polymorphisms reported were usually in terms of their variation from the Cambridge Reference Sequence (Anderson *et al.*, 1981, Andrews *et al.*, 1999). Until recently, mtDNA data from various populations have been compiled (Miller *et al.*, 1996; Budowle *et al.*, 1999). However, mtDNA data on Malaysian population is very limited. This had limited the application of mtDNA in forensic cases and study of mtDNA population genetics in Malaysia.

As a starting point, the Modern Malay, which is the largest population in Malaysia (55% of the country population) and Orang Asli, which is the aborigines of Peninsular

Malaysia, were chosen as the sample groups in this study. Other main groups were not chosen at this point, which include Indians and Chinese, since mtDNA data had already exists for these populations (Indians and Chinese from mainland India and China) that could be used.

3.1.1 Geography of Malaysia

Malaysia is situated on the South China Sea, in the centre of Southeast Asia, with geographic coordinates of 2 30 N, 112 30 E. The total land area is 328,550 sq km with the coastline of 2,068 km for West Malaysia and 2,607 km for East Malaysia. West Malaysia, which also known as Peninsular Malaysia, covering almost 132,000 sq km and comprise of 11 states and 1 federal territory. East Malaysia, which separated from West Malaysia by South China Sea, comprise of two larger states (Sabah and Sarawak) and 1 federal territory (Figure 3.1). With the location of one to seven degrees north of the equator, Malaysian weather is hot and humid all year round. The temperature is between 26 °C to 36 °C and humidity between 70% to 90% (Malaysian Meteorological Service).

Figure 3.1: Map showing the geographic location of East and West Malaysia within South East Asia and the international and state boundaries.

(Source: www.lib.utexas.edu/maps)

3.1.2 Modern Malays

According to the Malaysian Constitution, a Malay is a person who was born locally, habitually speaks Malay, follows Malay custom and professes Islam. The Malay race inhabits particularly Peninsular Malaysia and portions of adjacent islands of Southeast Asia, east coast of Sumatra, coast of Borneo and smaller islands between these areas. The Malays have been associated using anthropological studies to the people of north-western part of Yunnan, in China. The proto-Malays were seafaring people, probably from coastal Borneo who expanded into Sumatra and Peninsular Malaysia as a result of trading and seafaring activities. The present day Malays of Peninsular Malaysia and coast of the Malay Archipelago are a mixture of different groups, which form what is called the Modern Malay. They were descendants of the proto-Malay mixed with modern Indians, Thai, Arab and Chinese. The present Modern Malays are scattered throughout Peninsular Malaysia and most numerous on the east coast.

3.1.3 Orang Asli

The indigenous peoples of Peninsular Malaysia are known as Orang Asli. The name is a Malay term, which means the 'original people'. They numbered nearly 100,000 peoples in 2002 (data from Malaysian Department of Orang Asli Affairs) and divided into three main tribal groups, the Semang (Negrito), Senoi and Aboriginal Malays. Each tribe is further divided into 6 smaller groups (Table 3.1). Orang Asli are not a homogeneous group. Each has its own language and culture and also varies in their phenotype and traditional economy.

Table 3.1: Tribal groups and the corresponding sub groups of the Orang Asli population in Peninsular Malaysia (Malaysian Department of Orang Asli Affairs, July 2002).

Main Group	Sub Group	Population
Negrito	Kensiu	224
	Kintak	235
	Lanoh	359
	Jahai	1,049
	Mandriq	145
	Batiq	960
Senoi	Temiar	15,122
	Semai	26,049
	Semoq Beri	2,488
	Jahut	3,193
	Mahmeri	2,185
	Chewong	403
Aboriginal Malay	Temuan/Belandas	16,020
	Semelai	4,103
	Jakun	16,637
	Orang Kanaq	64
	Orang Kuala	2,492
	Orang Seletar	801
Total		92,529

The first wave of modern human migration into Peninsular Malaysia was believed to bring the Semang group, followed by the Senoi who arrived with the second wave about 4,000 to 8,000 YBP (Bellwood, 1985) and finally the Aboriginal Malays who arrived from island Southeast Asia (Carey, 1976; Fix, 1995; Rayner and Bulbeck, 2001). Most of the anthropological interest had focused on the Semang and Senoi since they showed distinct phenotypic different from the Modern Malays. Semang peoples are small in stature, muscular, had darker skin and curlier hair than other peninsular peoples.

The ancestry of Semang was traced to the people associated with the Hoabinhian archaeology, dated approximately about 4,000 to 10,000 years ago in Peninsular Malaysia (Bellwood, 1993). Populations that are phenotypically and culturally similar to the Semang also exist in Southern India, Sri Lanka, Indonesia, Papua New Guinea and China (Endicott *et al.*, 2003). Currently, the Semang inhibit mostly the mountainous area of Kelantan, Terengganu and northern regions of Perak, Kedah and Pahang. They reside in villages, which have been developed by the local authorities for Orang Asli all over Peninsular Malaysia. These villages are provided with electricity, water supply, school and other basic facilities in order to provide a better quality of living for Orang Asli.

3.2 Rationale of Samples Collection

Malaysia is a multiracial country with the Modern Malays as majority population. They make up about 55% of the total population of 22,662,365 peoples, as estimated on July 2002 (www.cia.gov/cia/publication/factbook/geos/my). Until recently, Malaysia is underrepresented of mtDNA data and therefore it is almost necessary to study mtDNA variations for Malaysian population, either for forensic applications or population study. The Modern Malays were chosen as the population sample since very limited mtDNA data is available for this particular population, as well as expected to have interesting genetic profiles. Since East Malaysia is separated fairly far from Peninsular Malaysia, with additional time and economy limitation, the Modern Malay sample collection was concentrated in Peninsular Malaysia.

The Kensiu and Jahai subgroups, from the Semang tribe, were the first two Orang Asli groups choose to be profiled. Since the Malays were believed not to be the descendants of the Orang Asli, investigation using mtDNA analysis might reveal some information regarding the relation of these two populations. The Semang, which believed as the first Orang Asli group to arrive in Peninsular Malaysia, was also thought to be related to Andaman islanders and aboriginals of the Philippines. Information from mtDNA analysis will help to generate more knowledge on this population, in addition to the application in forensic cases.

3.3 Samples Collection from Modern Malays

A total of 103 buccal cell samples were collected from the students and staffs of the Health Sciences School, University Sains of Malaysia. This campus is located in the Kelantan state and joined with a local Hospital. It was mostly occupied by medical students and staffs from all over Malaysia, particularly from Peninsular Malaysia. An official consent was obtained from the campus director prior to the sample collection. Informed consent was obtained from all donors. Each donor was given a short interview and requested to fill in an identification form in order to gain information about their maternal and paternal lineage. First and second-degree relatives of the donors were excluded from the sample group. Donors with mixed races background (according to the Malaysian law, if a Malay is married to other ethnics such as Chinese or Indian, their children are considered as Malays), which is a common scenario in Malaysia resulting from interethnic marriage, were also excluded from the sample group.

3.4 Samples Collection from Orang Asli

Orang Asli samples was collected from two Semang sub-groups, the Jahai and the Kinsiu. An official consent from the Malaysian Department of Orang Asli Affairs was obtained before collecting the samples. Consent was obtained from each donor. The Kinsiu samples were collected from an Orang Asli village in Baling, Kedah (west coast of Peninsular) and the Jahai samples were collected in Jeli, Kelantan (east coast of Peninsular). Donors with maternal or paternal relationship cannot be avoided since each village was occupied by a particular sub-group and had a very small percentage of

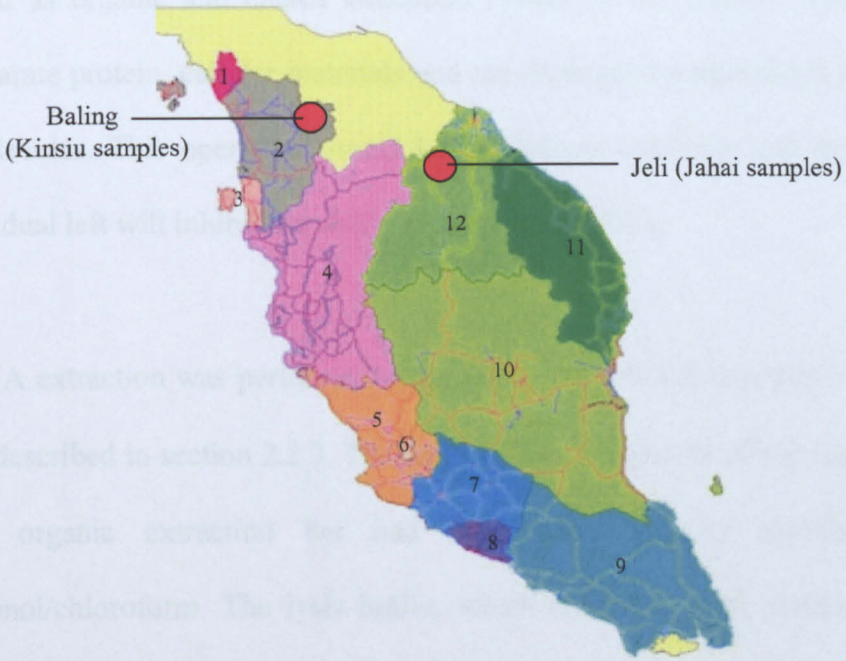
migrations either between tribes or groups. A total of 29 buccal cell samples were collected from the Kinsiu group and 30 samples from the Jahai.

3.5 Sample Handling and Storage

Handling collected samples in an appropriate way either from a crime scene or living human donors cannot be ignored. Carelessness can result in a specimen unfit for analyses. The most possible damage to happen was degradation of the DNA collected. Factors leading to the degradation of DNA include time, temperature, humidity (which leading to microorganisms growth), light (both sunlight and UV light) and exposure to various chemical substances (Rudin and Inman, 2002).

The hot and humid weather in Malaysia would promote relatively rapid DNA degradation if the swabs were left untreated. To avoid this, buccal swabs taken from donors were immersed into lysis buffer as soon as possible. An experiment carried out by Hadi, S.S. (2001) had showed that there was no marked degradation if buccal swabs were kept in lysis buffer, even after being exposed to high temperature. This method was therefore accepted and had been used for sample collection in Malaysia. Once buccal swabs were in the laboratory they were kept at -20 °C until ready for DNA extraction.

Figure 3.2: Map showing all states in Peninsular Malaysia which represented by the Modern Malay samples and the location of the Jahai and Kinsiu villages where the Orang Asli samples were collected.



States in Peninsular Malaysia:

1. Perlis
2. Kedah
3. Pulau Pinang
4. Perak
5. Selangor
6. Kuala Lumpur (Federal Territory)
7. Negeri Sembilan
8. Melaka
9. Johor
10. Pahang
11. Terengganu
12. Kelantan

3.6 DNA Extraction and Quantitation

DNA extraction is one of the vital step before any DNA sample can be tested, regardless of any typing procedure to be used. Several extraction methods have been developed, such as organic and chelex extraction (Walsh *et al.*, 1991). All methods serve to separate protein, cellular materials and non-biological materials (if present) from DNA molecules. This operation should be carried out carefully and thoroughly since any residual left will inhibit the ability to analyse the DNA.

DNA extraction was performed using Puregene® DNA Extraction kit (Gentra System), as described in section 2.2.3. The methodology applied by this extraction kit was based on organic extraction but had been simplified by omitting the usage of phenol/chloroform. The lysis buffer, which contained SDS, broke the cell walls and released DNA. Proteinase K was added to help breaking down the proteins which later being separated from DNA molecules by adding protein precipitation solution. Isopropanol was used to precipitate the DNA, which finally washed with 80% ethanol to remove any salt residues. DNA precipitate was resuspended in rehydrating solution containing EDTA to help better preservation. EDTA will bind any free magnesium that is needed in nuclease activity, therefore preventing the enzyme from digesting DNA. High Molecular Weight (HMW) DNA extracted from buccal swabs using the above method is shown in Figure 3.3.

The quantity of DNA extracted was measured using QuantiBlot® Human DNA Quantitation Kit before proceed with further analysis. This is to ensure optimal results obtained especially when dealing with PCR based analysis, which had a narrow

concentration range that worked best. The procedure is based on the hybridisation of a biotinylated oligonucleotide probe to DNA samples immobilised on a nylon membrane. The probe provided is complementary to a primate specific alpha satellite DNA sequence at locus D17Z1. Subsequent binding of the enzyme conjugate (horseradish peroxidase-streptavidin) to the biotin moiety of the probe allows for either colorimetric or chemiluminescent detection. Figure 3.4 showing the nylon membrane of colorimetric detection for the quantitated samples.

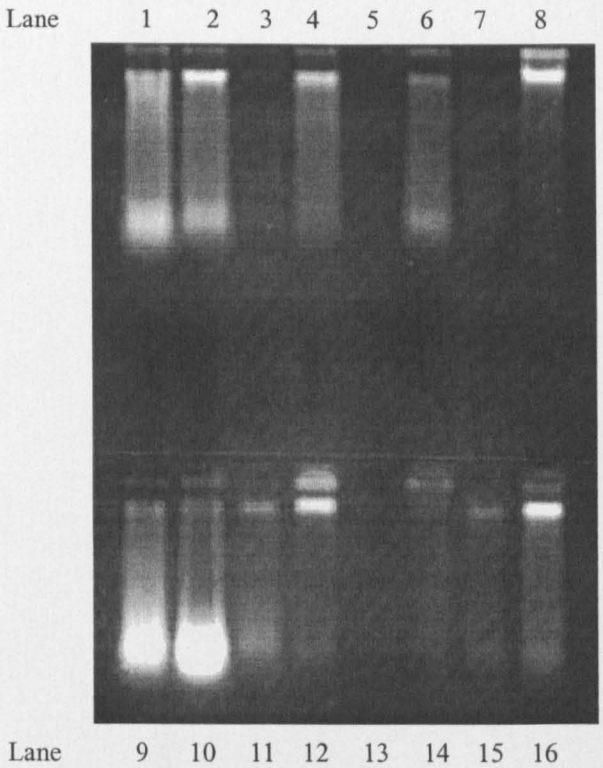


Figure 3.3: Agarose gel showing the High Molecular Weight DNA extracted from buccal swabs.

Lane 1-16: 4 µl of HMW DNA was loaded onto 1.2% agarose gel containing 1 µl ethidium bromide (10 mg/ml) and electrophoresed at 80 V for 45 min using 1X TBE buffer.

Figure 3.4: Nylon membrane showing blue dots with varies intensity after being developed in the colorimetric detection for DNA quantitation.

The quantity of each DNA sample was determined by comparison to the signal intensity of human DNA standard provided with the kit. A total of 5 µl of the diluted DNA standard was loaded on lane H6 to H12, along with DNA Calibration 1 and DNA Calibration 2 on lane C1 and G12, respectively. The rest of the lanes were loaded with 2 µl of the Modern Malay samples.

H6: DNA standard A, 2 ng/µl (10 ng)

H7: DNA standard B, 1 ng/µl (5 ng)

H8: DNA standard C 0.5 ng/µl (2.5 ng)

H9: DNA standard D 0.25 ng/µl (1.25 ng)

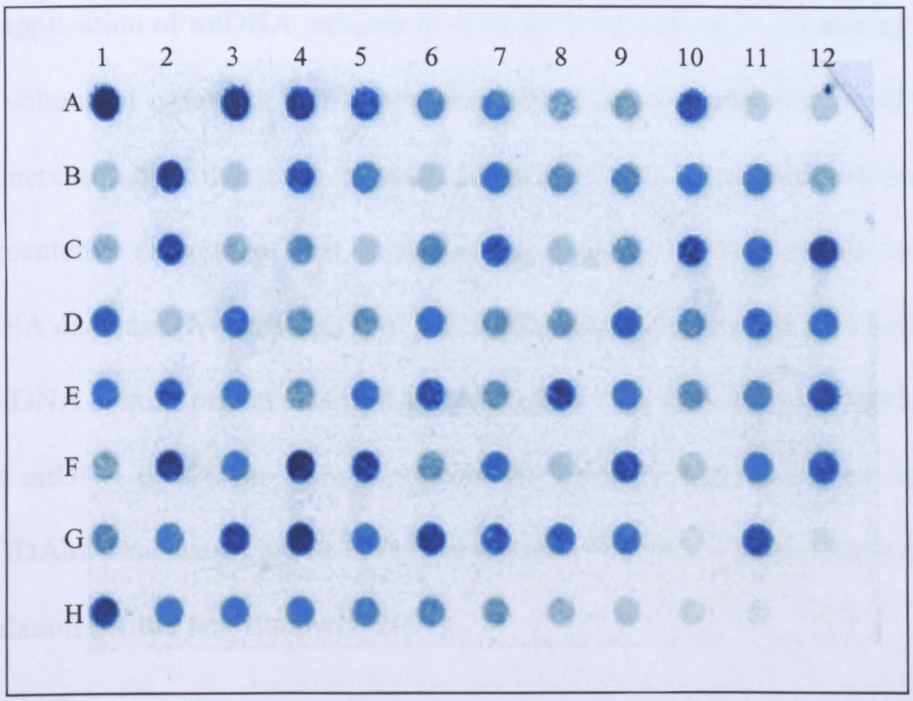
H10: DNA standard E 0.125 ng/µl (0.625 ng)

H11: DNA standard F 0.0625 ng/µl (0.3125 ng)

H12: DNA standard G 0.03125 ng/µl (0.15625 ng)

C1: DNA Calibration 1, 0.7 ng/µl (3.5 ng)

G12: DNA Calibration 2, 0.1 ng/µl (0.5 ng)


1 2 3 4 5 6 7 8 9 10 11 12
A
B
C
D
E
F
G
H

3.7 Discussion

The application of mtDNA analysis in forensic DNA testing is expanding very rapidly. The volume of casework and the number of laboratories performing mtDNA analysis had increased over the years. In order to perform mtDNA analysis of one population, representative samples of that particular population must be profiled for building up mtDNA database. A compilation of nucleotide substitution in the HVI and HVII region of mtDNA control region was first constructed in 1996 (Miller *et al.*, 1996). Since then, more mtDNA data from various populations had been developed. As reported in the SWGDAM database, only one HVI data and six HVII data were available for Malaysian population (Miller and Budowle, 2001).

As a starting point for developing mtDNA data, as well to study the pattern of mtDNA variations for Malaysian population, samples were collected from two populations of Peninsular Malaysia, the Modern Malays and Orang Asli. Sampling of the Modern Malays was done at University Sains Malaysia campus in Kubang Kerian, Kelantan. Since this campus was occupied by students and staffs from all over Malaysia, the samples collected was random enough with representatives from each state in Peninsular Malaysia. Each donor was shortly interviewed in order to establish his or her ancestry. The Malay samples were collected from individuals without mixed ethnic background for at least three generations.

Sampling of the Orang Asli population was done in two different locations within Peninsular Malaysia. The Kinsiu subgroup was collected in an Orang Asli village in the Baling district, in the Kedah state. This village was occupied by approximately 50

Kinsiu families. The Jahai subgroup was collected in another Orang Asli village in the Jeli district, in Kelantan state. This is a larger Orang Asli village, with approximately 150 families. Only 59 samples were managed to collect from both villages since many of the Orang Asli did not wish to participate as donors. Sample collection was also monitored by an officer from the Malaysian Department of Orang Asli Affair.

Buccal cells sampling was chosen as the sample collection method. Buccal epithelial cells were collected from the donors using cotton swabs. This sampling method was convenient, safe and donor friendly. The consent procedure was also much easier than obtaining blood sample. In addition, the buccal cells were also shown to be stable in extreme temperature, DNA yield was abundant and PCR failure rate was lower than DNA collected from blood (Richards, 1993). DNA yield from samples collected in this study were very good with the highest concentration of 50 ng/μl. This procedure is therefore suitable and satisfactory to be used for sample collection.

CHAPTER 4: SEQUENCING ANALYSIS OF THE MITOCHONDRIAL HYPERVARIABLE REGION 1

4.1 Introduction

Sequence variation within the hypervariable region I (HVI) and hypervariable region II (HVII) have been intensively studied. Polymorphisms occur across the 1.1 kb control region at relatively high frequencies, however, the HVI has consistently been shown to contain the highest density of polymorphisms (Richards and Macaulay, 2000; Meyer *et al.*, 1999; Lee *et al.*, 1997). Human mtDNA control region was also identified to have complex evolution pattern, with higher occurrence of nucleotide transitions than transversions, higher number of pyrimidine transitions in the L-strand than purine transitions and substitution rates which vary among sites (Tamura and Nei, 1993; Wakeley, 1993; Kocher and Wilson, 1991; Aquadro and Greenberg, 1983)

To date, thousands of individual HVI sequences identified from European, African and Asian population have been published (Imaizumi *et al.* 2002, Fucharoen *et al.*, 2001, Tsai *et al.* 2001; Wittig *et. al.*, 2000, Burckhardt *et al.*, 1999; Handt *et al.*, 1998 and Lee *et al.*, 1997). HVI sequence data, along with RFLP analysis of the coding region has been used in the study of human evolution, population genetics and also in forensic casework.

Sufficient reference data is vital before any useful forensic information can be obtained from mtDNA analysis. Malaysian population is underrepresented in the published mtDNA datasets. Sequencing of HVI region was a starting point to develop mtDNA

database for Malaysian population. The Modern Malay and Orang Asli, the biggest and the aborigine population of Peninsular Malaysia, respectively, were the first two groups chosen for this study. Besides the application in forensic casework, information gained from the HVI sequence data will help us to understand the diversity of these populations, as well the relationship to other populations studied before.

4.2 Sample

A total of 103 samples were collected from the Modern Malays and 59 samples from the Orang Asli in Peninsular Malaysia. mtDNA was extracted from buccal cells using Puregene® DNA Extraction kit (Gentra System). The HVI region was amplified, sequenced and analysed. Polymorphism was reported by aligning each individual sequence to the Cambridge Reference Sequence (Anderson *et al.*, 1981; Andrews *et al.*, 1999).

4.3 Amplification of HVI Region

PCR of HVI region for both Modern Malay and Orang Asli population was carried out to amplify a highly polymorphic fragment, ranging from position 16,001 to 16,400 in the mtDNA control region (Anderson *et al.*, 1981). The forward primers used in the PCR reaction (L15996 and L15926) were well established in HVI amplification and being used in other studies before (Mountain *et al.*, 1995; Baasner *et al.*, 1998; Szibor *et al.*, 2000). The reverse primer (H16431) was designed manually by referring to the CRS (Anderson *et al.*, 1981; Andrews *et al.*, 1999).

The PCR amplicon produced using L15996 and H16431 was 476 bp in length (including primers) (Figure 4.1) and PCR product amplified by L15926 and H16431 was 543 bp long (Figure 4.2). Each population had only one sample that failed to be amplified (sample numbered as M67 for Modern Malay and OA25 for Orang Asli). This might due to an error while collecting those particular samples (sufficient cells were possibly not collected on swabs) or an error during the extraction process. PCR was repeated several times for both samples but no product was observed after examination on an ethidium bromide stained agarose gel. M67 and OA25 were then excluded from the sample group, leaving a total of 102 and 58 samples for Modern Malay and Orang Asli, respectively.

Figure 4.1: Electrophoresis of PCR product for primer L15996 and H16431.

Lane 1: 1 kb DNA ladder (1 μg/μl), 1 μl loaded.

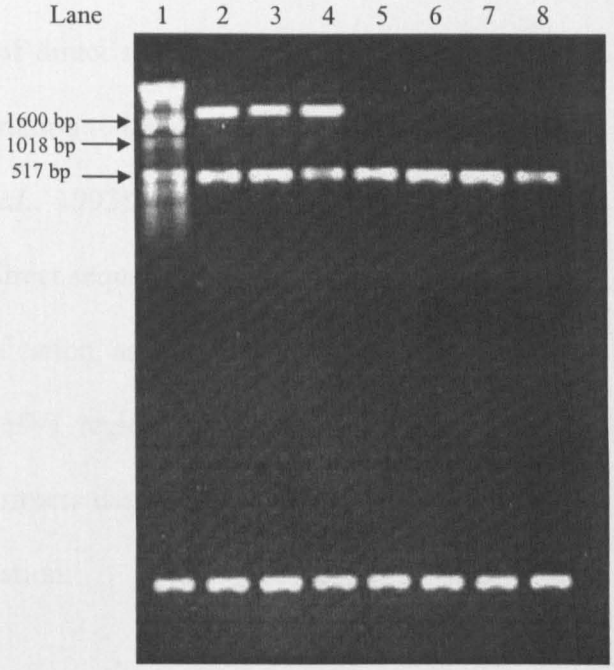
Lane 2-8: PCR product of HVI region, amplified using L15996 and H16431 primer (476 bp).

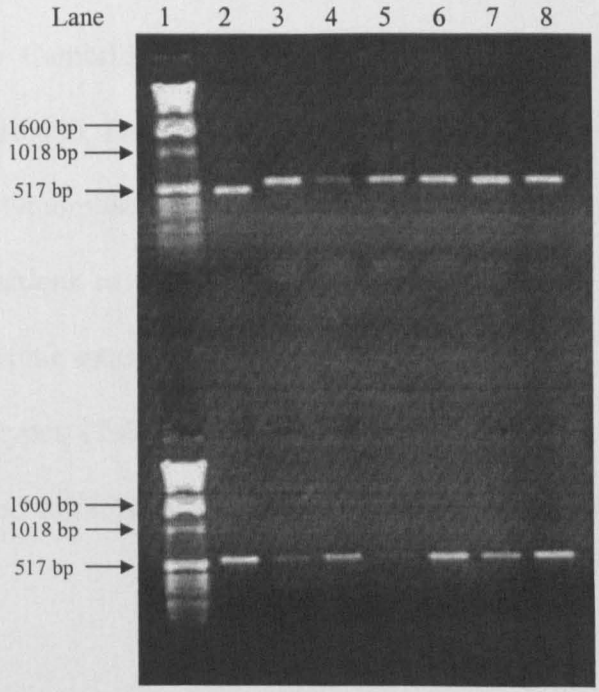
Figure 4.2: Electrophoresis of PCR product for primer L15926 and H16431.

Lane 1: 1 kb DNA ladder (1 μg/μl), 1 μl loaded.

Lane 3-8: PCR product of HVI region, amplified using L15926 and H16431 primer (543 bp).

*2 μl of the post-PCR reaction was loaded onto 1.2% agarose gel containing 1 μl EtBr (10 mg/ml) and electrophoresed at 80 V for 45 min using 1X TBE buffer.


Lane 1 2 3 4 5 6 7 8
1600 bp
1018 bp
517 bp


Lane 1 2 3 4 5 6 7 8
1600 bp
1018 bp
517 bp
1600 bp
1018 bp
517 bp

4.4 Sequencing of HVI Region

The availability of direct sequencing of PCR products had allowed fast and accurate sequencing information to be obtained from the mtDNA control region (Hopgood *et al.*, 1992; Piercy *et al.*, 1993). Point mutations within the D-loop region can be easily examined using direct sequencing of the amplified product. This had been a great utility in forensic identification, as well in population studies. In this study the amplified PCR fragment of the HVI region was analysed by automated sequencing as describe in section 2.2.5.3. Primers used in the sequencing reaction were the same as those used in the PCR amplification.

4.4.1 Sequence Analysis for the Modern Malay Samples

A total of 94 polymorphisms were recorded within the 102 sequences (Table 4.1) obtained for the Modern Malay population. Substitutions were detected by aligning each sequence with the Cambridge Reference Sequence (Anderson *et al.*, 1981) from position 16,001 to 16,400. The Gene Jockey software was used to facilitate the aligning process. All the polymorphisms that were observed were nucleotide substitutions. The ratio between transitions to transversions was 87:7, with two positions showed two variants of polymorphic sequence (shown in boldface in Table 4.1). In total there were 75 different haplotypes (Table 4.2). Only one sequence within the Modern Malay samples displayed no variation to the Cambridge Reference Sequence.

Table 4.1: Sequence polymorphisms and nucleotide frequency in HVI region from 102 Modern Malay samples. Each polymorphism was reported from variation to the Cambridge Reference Sequence (Anderson *et al.*, 1981; Andrews *et al.*, 1999).

Position	16051	16071	16086	16092	16093	16095	16108	16112	16124	16126	16129	16136	16140	16144	16145	16147	16148	16154	16157	16162	16166	16168	16169	16172	16176	16177	16179	16182	16183	16184	16185	16189	16191	16192	16193	16203	16204	16209	16223	16224	16231	16234	16235	16239	16241	16243	16249
CRS	A	C	T	C	T	C	C	T	T	G	T	T	T	G	C	C	T	T	A	A	C	C	T	C	A	C	A	A	A	C	C	T	C	C	C	A	T	T	C	T	T	C	A	C	A	T	T
M1																																								T								
M2															T															C	A		C					C					G				
M3										A		C																											T								
M4						T				A		C																											T								
M5												C																		C			C														
M6									C																																	C					
M7																																							T								
M9																																							T					T			
M10	G																																	C		T	C										
M11							T			A									G				C																								
M12					C																T									A								T								
M13					C																																T								
M14																																															
M15																												C	C		C					C											
M17												C																C	C		C					C											
M18																																						T								
M19			C							A																								C		T										
M20																							C								T				T											
M21																																					T										
M22								C				C																C	C		C															
M23																																					T										
M24	G		C													T															T					T										
M25										A													C																							
M26										A																					C	T				T										
M27			C																													T				T										
M28																																					T										
M29																																T															
M30*																																															
M31																						T															T										
M33							T			A									G				C																							
M34										A									G				C																							
M35										A																											T										
M36										A									G				C																							
M37			T			T				A									G				C																			T				
M38										A																					C	T				T										
M39																																					T										
M40																						G														T										
M41					C													C															T													
M42					C					A																										T										
M43															T													C		C				C					G					
M44												C																C		C															
M45											C																								T											
M46										A													C																							
M47												C																C		C															
M48										A													C																							
M49		T																					C							A		G														
M50																																				T											
M51										A									G				C																							
M52							T			A									G				C																							
M53																																				T											
M54												C																C		C															
M55																		T																													

Table 4.1: continued

Position	16255	16256	16257	16261	16263	16264	16265	16266	16269	16271	16272	16274	16278	16286	16287	16289	16290	16291	16292	16294	16295	16296	16297	16298	16299	16300	16304	16305	16309	16311	16319	16325	16327	16335	16344	16345	16355	16356	16357	16360	16362	16390	16398	16399
Anderson	G	C	C	C	T	C	A	C	A	T	A	G	C	C	C	A	C	C	C	C	C	C	T	T	A	A	T	A	A	T	G	T	C	A	A	C	C	C	T	T	C	T	G	G	A
M1																		T																										C	A		
M2				T																																									
M3							C			C																																			
M4							C			C																																			
M5								A																																		C			
M6																													C																
M7																				T																						C			
M9																G																									C				
M10					C							T																	A																
M11																						C																					A		
M12												T																																	
M13		T										A																																	
M14																				T	T				C															T					
M15												A																				G												
M17												A																				G												
M18																													C																
M19											G																																		
M20																																								T					
M21			A	T															T	T																									
M22				T				A																																			A		
M23																				T																						C			
M24																																										C	A		
M25																						C						C																
M26																							C																						
M27																							C																						
M28				T																T																									
M29																																													
M30*																																													
M31						T																						C																
M33																						C																						
M34																						C																						
M35																	T											C		C														
M36																						C																						
M37																					G	C																						
M38																					C																							
M39																																										A		
M40					C			T			A																	C				G										
M41																																												
M42		T							C																															C			
M43																																												
M44								A														C																						
M45			A	T															T	T																									
M46																						C						C																
M47								A																																				
M48																						C																						
M49																														C														
M50																												C																
M51																						C																						
M52																						C																						
M53				T																																						C	A		
M54				T				A	G																																				
M55																				T	T				C																				

* Sample showing sequence identical to the Cambridge Reference Sequence.

Table 4.1: continued

Position	160511	160786	160892	160935	160958	161024	161126	161369	161404	161445	161474	161548	161577	161622	161668	161689	161726	161779	161791	161812	161833	161845	161859	161892	161923	162001	162079	162173	162234	162314	162345	162359	162413	162433	162499
Anderson	A	C	T	C	T	C	C	T	T	G	T	T	G	C	C	T	T	A	A	C	C	T	C	A	C	A	A	A	C	C	T	C	C	C	A	T	T	C	T	T	C	A	C	A	T	T
M56									A									G			C														C				
M57									A		C			T																					T				G		
M58						T			A									G			C																			
M59																														C									
M60				C																										C		T	C						
M61																								C							T								
M62				C					A																						T				T				
M63									A					C							C	C		C						C									
M65																															T								
M66																															T								
M67																																							
M68									A											C																			
M69											C											C		C															
M70						T			A									G			C																		
M71																															T								
M72				C					A																						T								
M73												T									C	A		C						C				G				
M74										C											C			C														
M75				C					A																						T								
M76																																							
M77																															T								
M78																														C		T	C						
M79																															T								
M80																															T		T						
M81										C											C	C		C						C									
M82																															T								
M83																															T								
M84																															T								
M85									A		C		T		G												T				T				G		C	
M86																G																						
M87						T			A						G			C																				
M88				C																										C		T	C						
M89																																							
M90																															T		T						C
M91																															T		T						
M92																						C		C							T								
M93																			T																				
M94										C												C		C															
M95									A									C																				
M96			C						A																					C		T							
M97									A																	T					T								
M98												A																											
M99				C					A																						T								
M100																															T								
M101																																							
M101				C					A																						T								
M102																															T								
M103									A									C													T								
M104										C												C		C						C									
M105											C											C		C														C	
M106												A								G					T					T									
M107						T			A						G			C																					
freq	0.020	0.010	0.039	0.108	0.010	0.078	0.010	0.034	0.032	0.012	0.070	0.022	0.029	0.021	0.011	0.012	0.077	0.011	0.010	0.019	0.016	0.047	0.072	0.021	0.062	0.078	0.011	0.011	0.016	0.067	0.053	0.049	0.014	0.029	0.021	0.020	0.012	0.020

Table 4.1: continued

Position	16255	16256	16257	16261	16263	16264	16265	16266	16269	16271	16274	16278	16286	16287	16289	16290	16291	16292	16294	16295	16296	16297	16298	16299	16300	16304	16305	16309	16311	16319	16325	16327	16335	16344	16353	16355	16356	16360	16362	16390	16398	16399
Anderson	G	C	C	C	T	C	A	C	A	T	A	G	C	C	C	A	C	C	C	C	C	T	T	A	A	T	A	A	T	G	T	C	A	A	C	C	C	T	T	C	T	G	G	A
M56																										C																		
M57	A						C																		G				C					G										
M58																										C																		
M59																						C															T			C				
M60				C								T																		A														
M61																						C									T												G
M62													T		T														C											C				
M63			T																																									
M65																			T																					C				
M66			T																										C										C	A			
M67																																												
M68																										C																		
M69							A																																				
M70																										C			C														
M71																			T																					C				
M72		T							C																															C				
M73																																												
M74							A																																				
M75		T							C																															C				
M76																																												
M77																																												
M78				C								T																		A														
M79			T															T																						C	A			
M80			T												T																													
M81											A					T																												
M82																																										C	A		
M83									C					T																A						C				C			
M84			T																																						C	A		
M85						C																							C				G										
M86							T																			C			C														
M87																										C																		
M88				C								T																		A														
M89																																												
M90			T												T																										T			G
M91																									G				C														
M92																													C											C				
M93							T																															C						
M94							A																																					
M95																										C																		
M96										G																																		
M97																							C																					
M98							T																			C		G		C						C						
M99		T							C																																			
M100			A	T													T	T																										
M101																																												
M101		T							C																															C			
M102																			T																					C				
M103		T																								G	G																	
M104																																												
M105																																												
M106																T										C																		
M107																										C																		
freq	0.010	0.069	0.029	0.037	0.049	0.019	0.039	0.118	0.078	0.024	0.049	0.010	0.011	0.013	0.029	0.029	0.037	0.078	0.020	0.039	0.020	0.021	0.020	0.025	0.215	0.012	0.157	0.049	0.021	0.010	0.022	0.029	0.011	0.011	0.029	0.011	0.029	0.011	0.206	0.088	0.010	0.010

Table 4.2: List of 75 haplotypes, number of time observed and frequency of each haplotype in the Modern Malay population.

Haplotypes	Number of time observed	Haplotype frequency
1. 223	1	0.0098
2. 192	1	0.0098
3. 223, 311	2	0.0196
4. 223, 390	1	0.0098
5. 129, 172, 304	3	0.0294
6. 126, 231, 311	1	0.0098
7. 223, 261, 295	1	0.0098
8. 223, 295, 362	5	0.0490
9. 223, 362, 390	1	0.0098
10. 93, 157, 201	1	0.0098
11. 176, 266, 357	1	0.0098
12. 86, 192, 223, 297	1	0.0098
13. 93, 223, 256, 274	1	0.0098
14. 166, 266, 304, 311	1	0.0098
15. 129, 192, 223, 297	1	0.0098
16. 129, 172, 304, 311	2	0.0196
17. 129, 162, 172, 304	3	0.0294
18. 179, 223, 264, 311	1	0.0098
19. 168, 295, 296, 304	1	0.0098
20. 172, 193, 223, 344	1	0.0098
21. 223, 234, 300, 311	1	0.0098
22. 223, 261, 362, 390	2	0.0196
23. 223, 291, 362, 390	1	0.0098
24. 223, 239, 289, 356	1	0.0098
25. 223, 234, 261, 290	1	0.0098
26. 295, 296, 304, 354	1	0.0098
27. 209, 298, 355, 362	1	0.0098
28. 93, 129, 223, 256, 271	1	0.0098
29. 93, 129, 223, 256, 271, 362	4	0.0392
30. 93, 169, 184A, 223, 278	1	0.0098
31. 93, 209, 223, 224, 263, 278, 319	2	0.0196
32. 129, 140, 223, 265C, 271	1	0.0098
33. 129, 223, 290, 311, 325	1	0.0098
34. 129, 162, 172, 224, 304	1	0.0098
35. 129, 189, 192, 223, 297	2	0.0196
36. 129, 172, 223, 256, 305, 309	1	0.0098
37. 129, 144, 148, 223, 241, 255, 265C, 300, 311, 343	1	0.0098
38. 129, 144, 148, 162, 192, 223, 241, 249, 265C, 311, 343	1	0.0098
39. 145, 181, 192, 223, 291, 304	1	0.0098
40. 136, 223, 257A, 261, 292, 294	1	0.0098

41. 71, 172, 192A, 207, 325	1	0.0098
42. 145, 266, 304, 309, 325, 356	1	0.0098
43. 51, 86, 148, 185, 223, 362, 390	1	0.0098
44. 86, 129, 209, 223, 272	2	0.0196
45. 223, 257A, 261, 292, 294	2	0.0196
46. 223, 261, 311, 362, 390	1	0.0098
47. 223, 261, 294, 362, 390	1	0.0098
48. 223, 234, 249, 261, 290, 360, 399	1	0.0098
49. 223, 271, 287, 319, 356, 362	1	0.0098
50. 108, 129, 162, 172, 304, 311	1	0.0098
51. 108, 129, 162, 172, 304, 398	1	0.0098
52. 108, 129, 162, 172, 304	5	0.0490
53. 92, 108, 129, 162, 172, 234, 299,304	1	0.0098
54. 177, 223, 263, 266, 274, 311, 343	1	0.0098
55. 51, 209, 223, 224, 263, 278, 319	1	0.0098
56. 209, 223, 224, 263, 278, 319	1	0.0098
57. 95, 129, 140, 223, 265C, 271	1	0.0098
58. 93, 129, 223, 234, 286, 290, 311,362	1	0.0098
59. 147, 183C, 184A, 189, 217, 235, 261	1	0.0098
60. 140, 183C, 189, 266A	6	0.0588
61. 140, 183C, 189, 266A, 362	1	0.0098
62. 140, 183C, 189, 266A, 304	1	0.0098
63. 140, 183C, 189, 243, 311	1	0.0098
64. 140, 183C, 189, 261, 266A, 269	1	0.0098
65. 140, 182C, 183C, 189, 217, 274, 335	1	0.0098
66. 147, 183C, 189, 217, 235	1	0.0098
67. 147, 183C, 184A, 189, 217, 235	1	0.0098
68. 140, 182C, 183C, 189, 217, 274, 291	1	0.0098
69. 183C, 189, 223, 311, 362	1	0.0098
70. 136, 183C, 189, 217	1	0.0098
71. 189, 223, 298, 327, 399	1	0.0098
72. 129, 154, 182C, 183C, 189, 217, 261	1	0.0098
73. 124, 140, 182C, 183C, 189, 261, 266A, 390	1	0.0098
74. 182C, 183C, 189, 217, 274, 335	1	0.0098
75. CRS	1	0.0098

The most frequently observed nucleotide substitutions in the Modern Malay population were at np 16,223 (nucleotide transition from C to T), 16,129 (nucleotide transition from G to A), 16,304 (nucleotide transition from T to C), 16,362 (nucleotide transition from T to C), 16,189 (nucleotide transition from T to C) and 16,172 (nucleotide transition from T to C). Each polymorphism was presented in the population at the rate of 53.9%, 34.3%, 22.5%, 20.6%, 20.6% and 19.6%, respectively.

4.4.2 Sequence Analysis for the Orang Asli Samples

A significantly lower amount of polymorphisms were observed in the Orang Asli samples compared to the Modern Malays. Only 28 polymorphisms were counted in a total of 58 samples, which were mostly nucleotide transitions. The ratio between transitions to transversions was 25:3. All polymorphisms noted were recorded along with the Cambridge Reference Sequence in Table 4.3. Only one position (np 16,240) showed two variants of polymorphic sequence. The highest frequency of nucleotide substitution was observed at np 16,295 (nucleotide transition from C to T), with 50% of the samples bearing this polymorphism. Nucleotide substitution at 16,086, 16,093, 16,129, 16,168, 16,223, 16,256, 16,271, 16,296, 16,304 and 16,362 were also identified with high frequency in the Orang Asli population. Each frequency is shown in Table 4.3. A total of 13 haplotypes were established for the Orang Asli population and 3 of them appeared very frequently within the sample population (Table 4.4).

Table 4.3: Sequence polymorphisms and nucleotide frequency in HVI region of 58 Orang Asli samples. Each polymorphism was reported from variation to the Cambridge Reference Sequence (Anderson *et al.*, 1981; Andrews *et al.*, 1999).

Position	16086	16093	16129	16140	16168	16183	16189	16209	16210	16223	16224	16240	16256	16257	16261	16263	16266	16271	16278	16292	16294	16295	16296	16304	16311	16325	16362	16381
Anderson	T	T	G	T	C	A	T	T	A	C	T	A	C	C	C	T	C	T	C	C	C	C	C	T	T	T	T	T
OA1			A							T			T					C									C	
OA2					T																	T	T	C				
OA3					T																	T	T	C				
OA4		C	A							T			T					C									C	
OA5		C	A							T			T					C									C	
OA6				C		C	C					C					A											
OA7		C	A							T			T					C									C	
OA8					T																	T	T	C				
OA9		C	A							T	C		T					C									C	
OA10				C		C	C										A											
OA11				C		C	C					C																
OA12		C	A							T			T					C									C	
OA13			A							T			T					C									C	
OA14					T																	T	T	C				
OA15		C	A					C		T					T				T							C		
OA16		C	A							T			T					C									C	
OA17		C	A							T		G	T					C									C	
OA18		C	A							T			T					C									C	
OA19					T																	T	T	C				
OA20					T																	T	T	C				
OA21		C	A					C		T					T				T							C		
OA22					T																	T	T	C				
OA23		C	A							T			T					C									C	
OA24				C		C	C					C																
OA26									G	T												T			C		C	
OA27		C	A							T		G	T					C									C	
OA28		C	A							T			T					C									C	
OA29					T																	T	T	C				
OA30		C	A							T			T					C									C	
OA31		C	A							T						C												C
OA32										T				A	T					T	T							
OA33	C				T																	T	T	C				
OA34		C	A							T			T					C									C	
OA35	C				T																	T	T	C				
OA36		C	A							T			T					C									C	
OA37		C	A							T			T					C									C	
OA38	C				T																	T	T	C				
OA39	C				T																	T	T	C				
OA40	C				T																	T	T	C				
OA41	C				T																	T	T	C				
OA42	C				T																	T	T	C				
OA43		C	A							T			T					C									C	
OA44	C				T																	T	T	C				
OA45	C				T																	T	T	C				
OA47	C				T																	T	T	C				
OA48					T																	T	T	C				
OA49	C				T																	T	T	C				
OA50		C	A							T			T					C									C	
OA51		C	A							T			T					C									C	
OA52										T				A	T					T	T							
OA53	C				T																	T	T	C				
OA54	C				T																	T	T	C				
OA55	C				T																	T	T	C				
OA56	C				T																	T	T	C				
OA57	C				T																	T	T	C				
OA58	C		A		T																	T	T	C				
OA59	C				T																	T	T	C				
OA60	C				T																	T	T	C				
Freq	0.328	0.362	0.414	0.069	0.483	0.069	0.069	0.034	0.017	0.448	0.017	0.086	0.345	0.034	0.069	0.017	0.034	0.345	0.034	0.034	0.034	0.500	0.483	0.483	0.017	0.034	0.362	0.017

Table 4.4: List of 13 mtDNA haplotypes, number of time observed and frequency of each haplotype in the Orang Asli population.

Haplotypes	Number of time observed	Haplotype frequency
1. 86, 168, 295, 296, 304	18	0.310
2. 93, 129, 223, 256, 271, 362	15	0.259
3. 93, 129, 223, 263, 381	1	0.017
4. 93, 129, 223, 224, 256, 271, 363	1	0.017
5. 93, 129, 209, 223, 261, 278, 325	2	0.035
6. 93, 129, 223, 240, 256, 271, 362	2	0.035
7. 129, 223, 256, 271, 362	2	0.035
8. 140, 183C, 189, 243T/C*, 266A	3	0.052
9. 140, 183C, 189, 266A	1	0.017
10. 168, 295, 296, 304	9	0.155
11. 210, 223, 295, 311, 362	1	0.017
12. 223, 257, 261, 292, 294	2	0.035
13. 86, 129G/A*, 168, 295, 296, 304	1	0.017

*Point mutation heteroplasmy

4.5 Statistical Analysis

The pattern of inheritance had made statistical analysis of mtDNA type much easier than any other genetic marker. Since mtDNA presented in each human being as haploid, determination of mtDNA type did not require the prerequisite of Hardy-Weinberg equilibrium for statistical analysis. The frequency of each mtDNA type within a population was easily determined by counting method, where the number of time a particular sequence being observed, divided by the number of sequence contained within the related database. Genetic diversity (h) calculated for Modern Malay population was 0.989 and the probability of two randomly selected individuals from this population to have identical mtDNA types (P) was 0.0202. Based on the Jahai and Kinsiu subgroup, genetic diversity calculated for Orang Asli was 0.818 and the probability of random match of two individuals (P) in this population was 0.1962. Calculation was done based on the following formulae (Tajima, 1989):

$$h = \frac{(1-\Sigma x^2)n}{n-1} \qquad P = \Sigma x^2$$

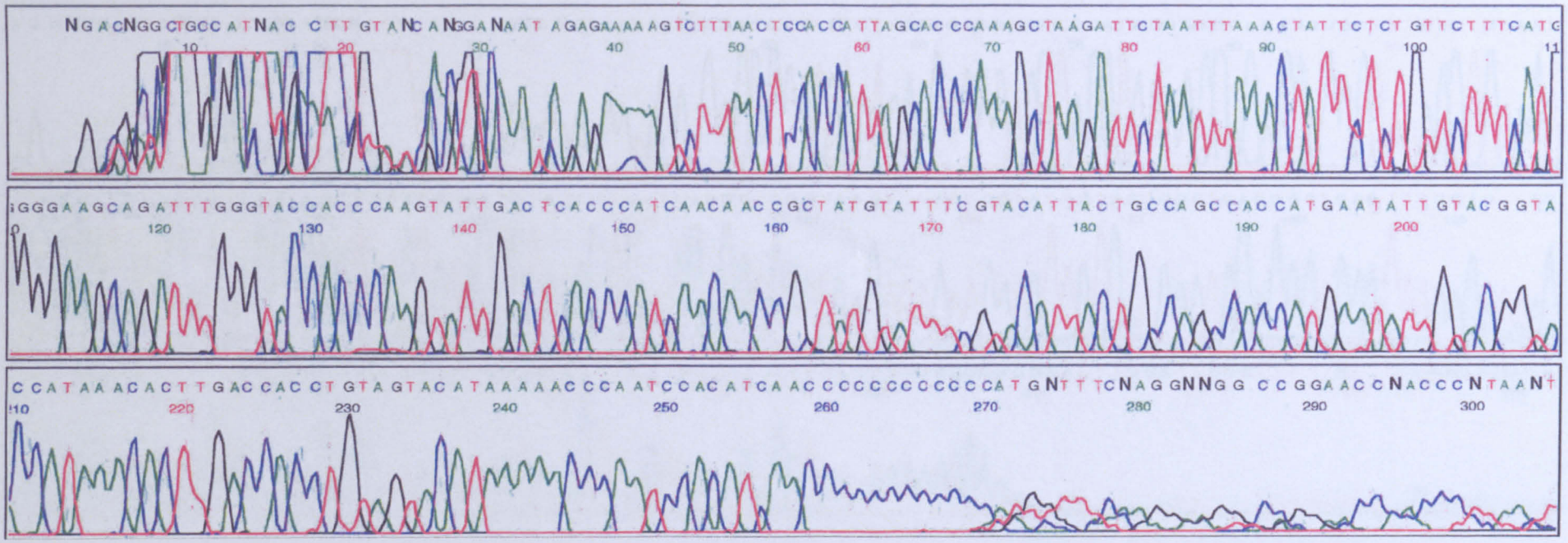
n= sample size x= frequency of each mtDNA type

4.6 C-stretch Region

In both population samples, a homopolymeric region containing serially repeated C residues was observed between positions 16,184 to 16,193 when a nucleotide transition from T to C occurred at np 16,189. Sequencing of these samples were quite complicated since the quality of the sequence had dramatically reduced downstream of this homopolymeric region. This was due to template molecules that were out of register (Parson *et al.*, 1998). Samples carrying this nucleotide transition were sequenced for both light and heavy strand using L15926 and H16431 primer, respectively. Sequencing for both strands had provided good reading upstream of the homopolymeric region (Figure 4.3 and 4.4). Sequencing of the light strand provided data between position 16,001 to 16,189 while the heavy strand provided data between position 16,189 to 16,400. Sequence data alignment for sample M17 (one of the sample with the 16,189 T to C transition) to the Cambridge Reference Sequence is shown in Table 4.5 and Table 4.6. Polymorphisms were located at np 16,140, 16,182, 16,183, 16,189, 16,217, 16,274 and 16,335 within the amplified fragment.

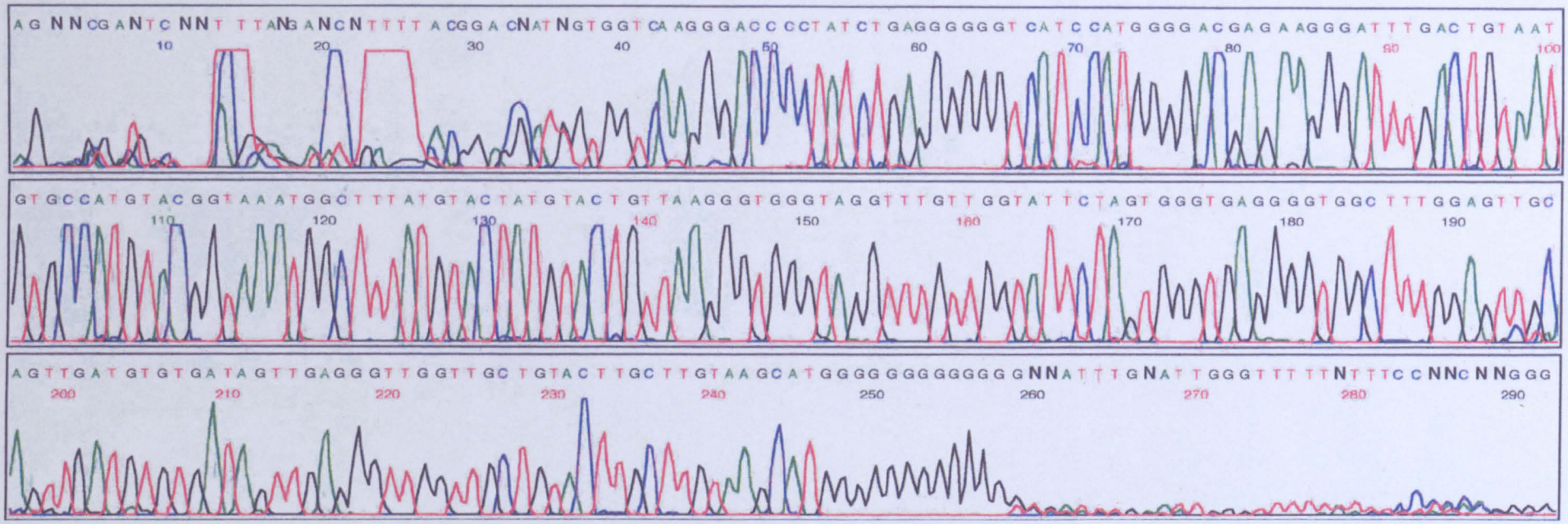
The longest C-stretch region observed within the sample populations contained 12 unbroken series of C residues, created by additional nucleotide transversion from A to C at np 16,182 and 16,183. Sequence variation from np 16,182 to 16,193 for the Modern Malay and Orang Asli samples bearing the 16,189 T to C transition is shown in Table 4.7.

Figure 4.3: Electropherogram showing the sequencing result of the amplified HVI fragment for sample M17 using primer L15926.



Note that the sequence quality had dramatically reduced beyond the homopolymeric region. Sequence quality upstream the homopolymeric region (from position 50 in the electropherogram) was good without any background interference. Sequencing using the light strand primer had provided reading between position 16,001 to 16,189.

Figure 4.4: Electropherogram showing sequencing result of the amplified HVI fragment for sample M17 using primer H16431.



Sequence quality was again reduced beyond the homopolymeric region. Sequencing using the heavy strand primer had provided unambiguous data from np 16,189 to 16,400.

Table 4.5: Sequence alignment of M17 light strand to Cambridge Reference Sequence. The top sequence represents the Cambridge Reference Sequence while the bottom sequence belongs to sample M17. Nucleotide transition from T to C was noted at np 16,140 and 16,189. Nucleotide transversion from A to C was noted at np 16,182 and 16,183, making a total of 12 cytosine residues within the homopolymeric region.

Table 4.6: Sequence alignment of M17 heavy strand to Cambridge Reference Sequence. The top sequence and bottom sequence belongs to Cambridge Reference Sequence and M17, respectively. Unambiguous sequence data started from position 16,189 upwards. Nucleotide transition from T to C, G to A and A to G was observed at np 16,217, 16,274 and 16,335, respectively.

Table 4.5:

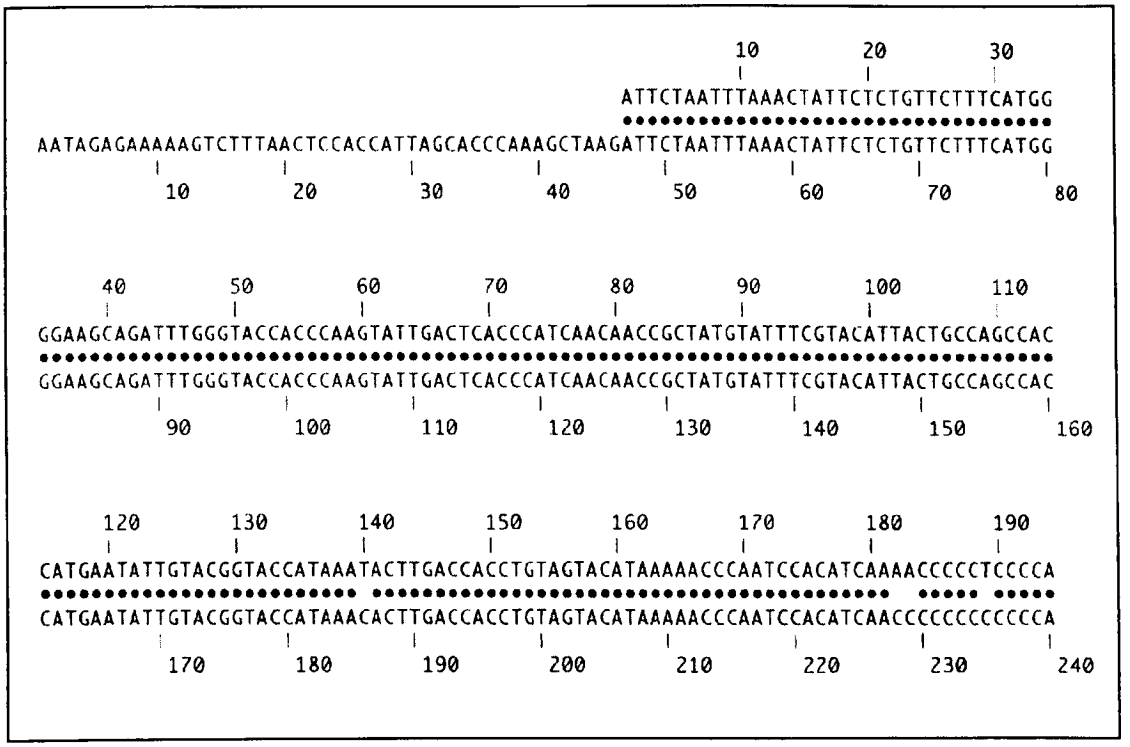


Table 4.6:

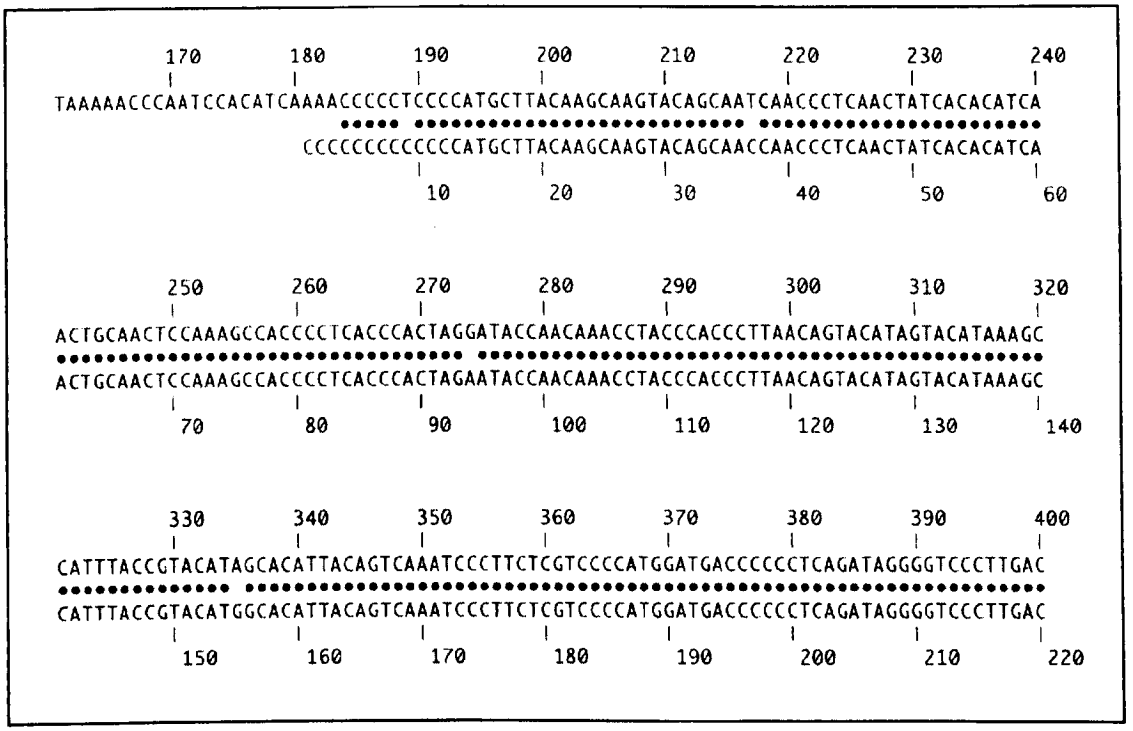


Table 4.7: Sequence variation between position 16,182 to 16,193 for samples bearing T to C transition at np 16,189 in the HVI region.

Sequence variation	Number of time observed	
	Modern Malay	Orang Asli
ACACCCCCCCCC	2	-
AACCCCCCCCCC	1	-
AACCCCCCCCTC	2	-
ACCCCCCCCCCC	10	4
CCCCCCCCCCCC	5	-

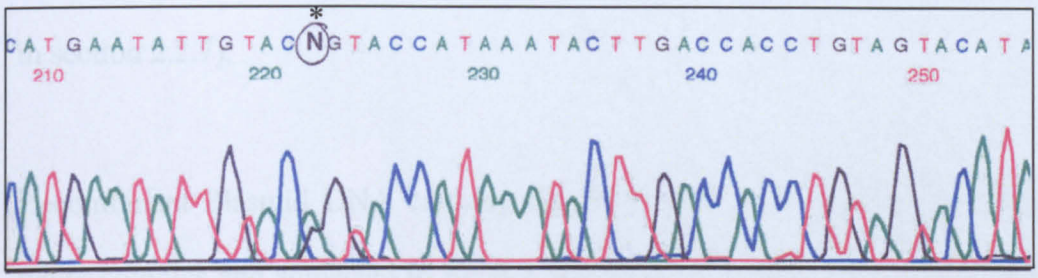
4.7 Point Mutation Heteroplasmy

The occurrence of heteroplasmy in mtDNA typing does not invalidate the use of mtDNA in forensic casework. Even though heteroplasmy sometimes can complicate data interpretation, this special feature had proven to improve the probability of a match, as seen in the Romanov case (Ivanov *et al.*, 1996). The existence of heteroplasmy could be confirmed if two bases were clearly observed above the background level of the sequence data in both strands of mtDNA (Carracedo *et al.*, 2000).

One point mutation heteroplasmy was detected in the Orang Asli sample (OA58) by direct sequencing of the amplified HVI region fragment. It was located at np 16,129, bearing both nucleotide G and A at this particular position. Careful examination of both light and heavy strand was carried out in order to confirm the existence of this heteroplasmic point mutation. Sequencing of this sample was repeated twice for both strands and a consistent mixed signal was observed without any ambiguities caused by background disturbance (Figure 4.5 and 4.6).

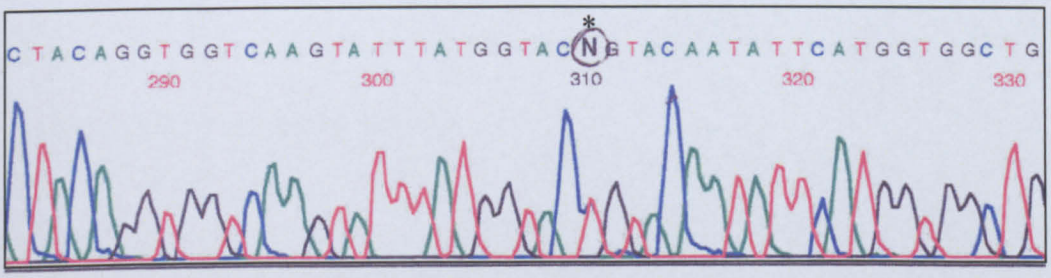
A second point mutation heteroplasmy was also detected in the Orang Asli sample, located at np 16,243. Three samples within the Kinsiu tribe, OA6, OA11 and OA24 carried both nucleotide T and C at this specific point. This heteroplasmic point mutation was however not able to be detected by direct sequencing of the PCR fragment. Sequence quality surrounding the heteroplasmic point mutation region was somehow very poor and the existence of the 16,189 T to C transition within these samples had made the condition worst. In order to resolve the sequence ambiguities for these three

Figure 4.5: Electropherogram showing sequencing of the HVI region light strand for sample OA58.



Point mutation heteroplasmy was observed at np 16,129* for sample OA58. The mixture of peak G and A at this position can be clearly seen.

Figure 4.6: Electropherogram showing sequencing of the HVI region heavy strand for sample OA58.



Point mutation heteroplasmy was again observed at np 16,129* for sample OA58. Both peak C and T at this position can be clearly seen.

samples, PCR fragments of the HVI region were cloned into pGEM®-T vector. These ligated products were then used to transform JM109 competent cells and the transformants were screened by the normal blue/white screening procedure (explained in section 2.2.7).

Recombinant Plasmid DNA carrying the HVI insert was isolated from the selected white colonies and subjected to *Eco*R I digestion for confirmation of the correct insert (Figure 4.7). Sequencing of this recombinant plasmid was done using M13 forward and reverse primer. Both sequencing direction had successfully resolved the sequence ambiguities for sample OA6, OA11 and OA24. Since each recombinant plasmid only carried one type of the HVI sequence, the point mutation heteroplasmy was clearly observed between different clones. The heteroplasmic point mutation detected from sequencing of the recombinant plasmids is shown in Figure 4.8.

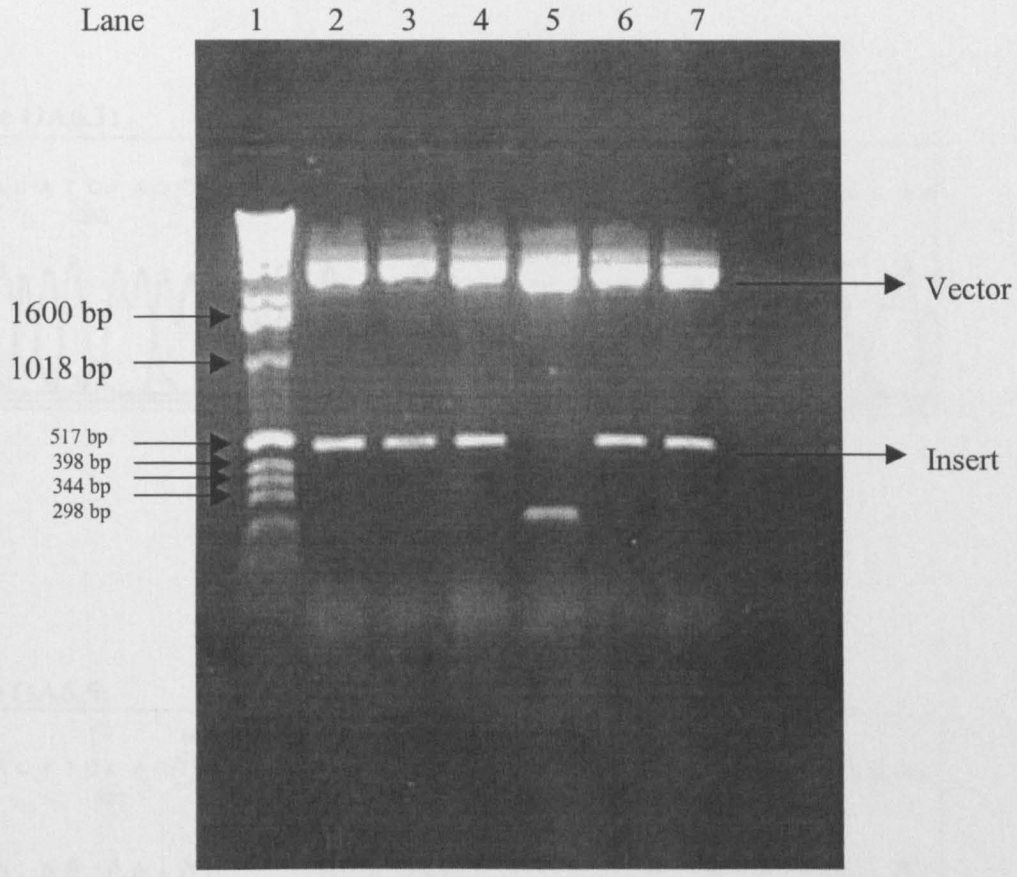


Figure 4.7: Electrophoresis of the digested plasmid isolated from white colonies using *Eco*R I restriction enzyme.

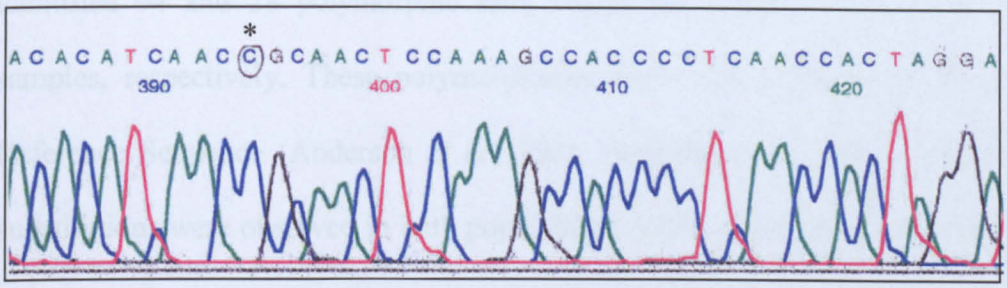
Lane 1: 1 kb DNA ladder, 1 µg/µl (1 µl loaded)

Lane 2-7: Recombinant plasmid digested with *Eco*R I.

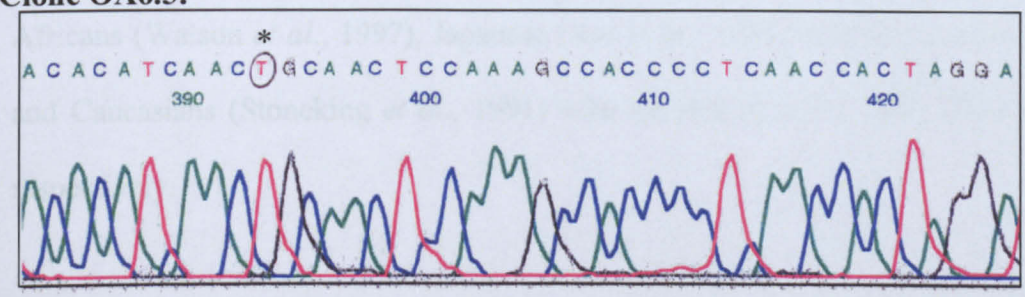
Recombinant plasmids isolated from the selected white colonies were screened for the correct insert by *Eco*R I digestion. 4 µl of the digested product was electrophoresed on 1.2% agarose gel for 50 min at 80 V using 1X TBE buffer. All selected clones were found bearing the HVI fragment, determined by the size of the insert (approximately 500 bp) except for clone on lane 5. The size of the insert for this clone was too small, most probably was the primer.

Figure 4.8: Point mutation heteroplasmy detected by sequencing of the recombinant plasmid DNA bearing the HVI region insert.

Clone OA6.1:



Clone OA6.5:



Electropherogram showing sequencing of clone OA6.1 and OA6.5 using M13 reverse primer. Clone OA6.1 carried a C at np 16,243 while OA6.5 carried a T at the same nucleotide position (shown with the asterisk sign in both electropherograms) within the HVI fragment. Both bases were clearly observed above the background level without any ambiguities.

4.8 Discussion

Analysis of the amplified HVI fragment by direct sequencing had successfully identified 94 and 28 polymorphic sites within the Modern Malay and Orang Asli samples, respectively. These polymorphisms were with reference to the Cambridge Reference Sequence (Anderson *et al.*, 1981, Andrews *et al.*, 1999). Only nucleotide substitutions were observed in both populations, which mostly involved transition. C to T transition at np 16,223 was predominant in the Modern Malay population with the frequency of 53.92%. The Orang Asli population also exhibits high frequency of this polymorphism (44.83%). The high frequency of this nucleotide transition in both populations was not very intriguing since it was also found in high frequencies in Africans (Watson *et al.*, 1997), Japanese (Seo *et al.*, 1998), Koreans (Lee *et al.*, 1997) and Caucasians (Stoneking *et al.*, 1991) with the rate of 92%, 72%, 78% and 44.2% respectively.

The most significant difference of the polymorphism frequency between the Modern Malay and Orang Asli population was observed at np 16,168 and 16,296. Nucleotide transition from C to T at both 16,168 and 16,296 positions occurred at high frequency (48.3%) in the Orang Asli population but was only found at the rate of 0.98% and 1.96%, respectively in the Modern Malay population. Both polymorphisms were observed at higher frequency in the Jahai subgroup (66.67%) than the Kinsiu subgroup (27.59%) and occurred simultaneously each time. Nucleotide transitions at 16,168 and 16,296 had never been reported at high frequency elsewhere. Other polymorphisms observed at high frequency in the Orang Asli compared to the Modern Malay were at np 16,093, 16,256, 16,271 and 16,295. In contrast to the Orang Asli population, the degree

of polymorphism observed within the Modern Malays was much evenly distributed between the samples. Besides the nucleotide transition at 16,223, only positions 16,129, 16,304 and 16,362 showed a high frequency of nucleotide transition within the population (34.31%, 22.55% and 20.59%, respectively).

A total of 75 and 13 haplotypes were attained for the Modern Malay and Orang Asli population, respectively, by using the information from the HVI sequence variation. In the Modern Malay population, 62 haplotypes (82.67%) were seen only once and 13 haplotypes (17.33%) had occurred more than once. Of the 13 haplotypes seen in multiple individuals, 7 were shared by 2 individuals, 2 were shared by 3 individuals, 2 were shared by 5 individuals, 1 was shared by 4 individuals and 1 was shared by 6 individuals. The small number of shared mtDNA type compared to the larger number of rare mtDNA type within the Modern Malay population was the pattern of mtDNA population structure observed elsewhere (Budowle *et al.*, 1999; Fucharoen *et al.*, 2001; Imaizumi *et al.*, 2002). The frequency of unique mtDNA sequence for the Modern Malay population can be improved in future by increasing the number of samples in the database and adding information from the HVII region sequence variation. The single observation of mtDNA type for HVI and HVII region ranged from 60.3% to 85.3% in the available database (Budowle *et al.*, 1999b). Since this study had only examined the HVI sequence variation, the total of 82.67% of unique mtDNA type offered a high mtDNA variation within the Modern Malay population.

In the Orang Asli population, the amount of unique mtDNA type observed was notably lower than the Modern Malays. Of 13 haplotypes observed, only 5 (38.46%) were not shared between the individuals. One haplotype was shared by as many as 18

individuals, 1 was shared by 15 individuals, 1 was shared by 9 individuals, 1 was shared by 3 individuals and 4 were shared by 2 individuals. The pattern of mtDNA variation within the Orang Asli population showed the small genetic pool for this population. The nearest explanation for the reduced variability of the Orang Asli population might be a result from the small population size and sampling from many maternally related individuals. Sequence analysis for the Orang Asli population had also shown limited intermixture between both subgroups. This can be clearly seen from the distribution of mtDNA type between the Jahai and Kinsiu subgroup with only two haplotypes were shared by both populations.

Two point mutation heteroplasmy at np 16,129 and 16,243 were observed within the Orang Asli population. The 16,129 heteroplasmic site was reported as a hot spot for heteroplasmy by DGGE analysis (Holland and Parsons, 1999). Sample OA58 from the Jahai subgroup was detected to carry both nucleotide G and A at np 16,129 by direct sequencing of the amplified HVI fragment. This heteroplasmic point mutation was easily detected since the sequence quality was very good, without any disturbance from the background artefact. Point mutation heteroplasmy might not be easily detected if high background signal exist in the sequence electropherogram. This imperative information might be overlooked and lead to wrong data interpretation. It is therefore very important to keep a high quality sequence when dealing with mtDNA analysis. Point mutation heteroplasmy was usually obvious when the ratio of peak heights was at least 1:4 (Melton and Nelson, 2001).

Another point mutation heteroplasmy was found at np 16,243 in sample OA6, OA11 and OA24 from the Kinsiu subgroup. Heteroplasmy at this position is rarely found in

previous studies (Budowle *et al.*, 2002). This heteroplasmic point could not be detected by direct sequencing of the PCR fragment since the sequence quality surrounding this heteroplasmic point was very poor. This might due to the additional nucleotide transition at np 16,189, creating a homopolymeric region which usually associated to the sequence quality drop, very near to this heteroplasmic site. The sequence ambiguities was resolved by cloning the HVI fragment into plasmid vector and sequenced using the M13 reverse and forward primer. Even though this method is very reliable in producing good sequence quality, it was not very practical for forensic application due to the time consumed and tedious procedure. Sequencing using junction primers complimentary to the junction between the homopolymeric region and the 5′ part of the template strand had also successfully resolve sequence ambiguities related to the homopolymeric region (Rasmussen *et al.*, 2002). This method is simpler and faster compared to the molecular cloning and more appropriate to be used in forensic casework.

Previous study had shown that heteroplasmy within the control region is rare and was not a result of somatic accumulation of sequence variation (Lagerstrom-Fermer *et al.*, 2001). Once inherited from the mother, heteroplasmy in the mtDNA control region was suggested to remain stable throughout life and did not undergo selection for either of the variants (Lagerstrom-Fermer *et al.*, 2001). Based on these suggestions, sample OA6, OA11 and OA24, which carried the same point mutation heteroplasmy at np 16,243 and also exhibited identical HVI sequence, may well be maternally related. This could be further examined and confirmed by STR analysis.

Sequence information of the HVI region had offered great potential of human identification in forensic science. The availability of direct sequencing of PCR product had allowed fast and reliable sequence information to be obtained from mtDNA control region. The occurrence of heteroplasmy within the control region and their stable inheritance mode had improved mtDNA utility in human identification cases since they add higher matching probabilities than relying on the polymorphic sequence variation only. Anyhow, application of mtDNA sequence variations to forensic casework is much more appropriate in identification of human remains such as from mass disaster, accident and war, rather than identification of criminals. Barring mutation, this limitation was due to identical mtDNA type existed between maternally related individuals.

CHAPTER 5: RESTRICTION FRAGMENT LENGTH POLYMORPHISM ANALYSIS OF MITOCHONDRIAL DNA CODING REGION.

5.1 Introduction

Restriction Fragment Length Polymorphism (RFLP) and direct sequencing analysis has been used extensively for the past 15 years to study the pattern of mtDNA variation. Sequence variation within mtDNA coding region has been primarily detected by using RFLP analysis (Horai *et al.*, 1984; Cann *et al.*, 1987; Harihara *et al.*, 1988; Stoneking *et al.*, 1990; Ballinger *et al.*, 1992; Torroni *et al.*, 1994; Kivisild *et al.*, 1999). This method has successfully revealed a number of stable polymorphic sites, which used to defined mtDNA haplogroups. Most of these established haplogroups were shown to be continent specific (Wallace, 1994). The mtDNA haplogroups network for the world populations are shown in Figure 5.1. Three of the most frequently observed mtDNA haplogroups in Southeast Asian populations are shown in Table 5.1.

Determination of mtDNA genotype has also utilised the sequence variation of the hypervariable region. Since good correlation between RFLP data and HVI sequence data has been established (Torroni *et al.*, 1996,1998; Macaulay *et al.*, 1999), classification of mtDNA genotypes should be more specific and precise when data from both methods were applied.

Figure 5.1: A skeleton network showing the world mtDNA haplogroups and their geographical location (Richards and Macaulay, 2000). Diagnostic mutations, mostly from HVI region and the coding region polymorphisms (responsible for creating RFLP sites) are indicated on the branches. Mutations shown are transitions (C to T or A to G), unless a transversion is indicated with a 't' suffix. Recurrent mutations are shown underlined (Richards and Macaulay, 2000). The numbering indicating base pair position relative to the Cambridge Reference Sequence (Anderson *et al.*, 1981; Andrews *et al.*, 1999). The tree was constructed by using information, in particular from Macaulay *et al.*, 1999, Quintana-Murci *et al.*, 1999 and Watson *et al.*, 1997.

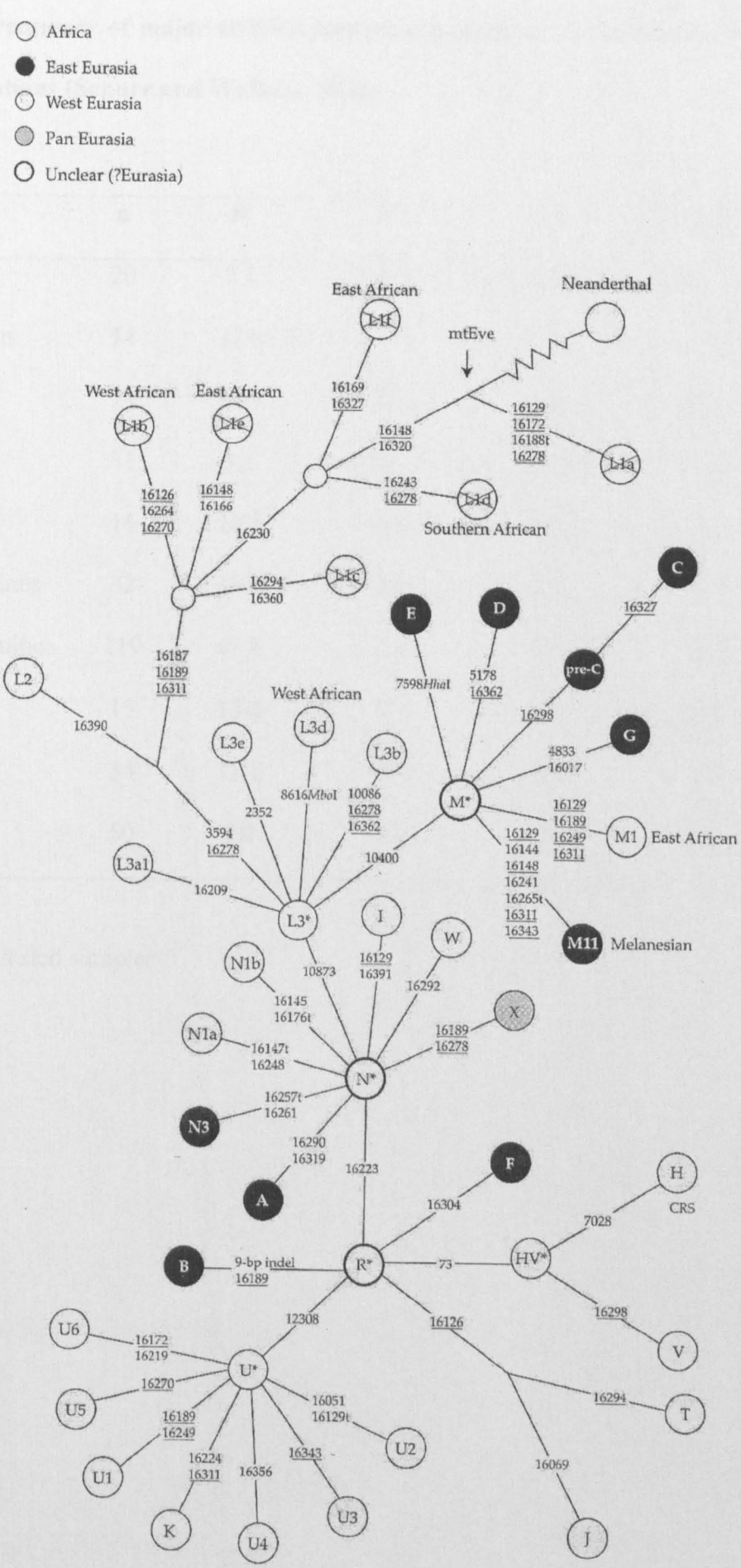

Africa
East Eurasia
West Eurasia
Pan Eurasia
Unclear (?Eurasia)
East African
L1f
Neanderthal
mtEve
West African
East African
L1b
L1e
16169
16327
16148
16320
16129
16172
16188t
16278
L1a
16126
16264
16270
16148
16166
16243
16278
L1d
Southern African
16230
16294
16360
L1c
C
16327
E
D
16187
16189
16311
7598HhaI
5178
16362
pre-C
16298
L2
16390
West African
L3d
L3e
L3b
G
4833
16017
8616MboI
10086
16278
16362
2352
M*
16129
16189
16249
16311
M1 East African
3594
16278
L3a1
16209
10400
16129
16144
16148
16241
16265t
16311
16343
M11 Melanesian
L3*
I
W
N1b
10873
16129
16391
16292
16145
16176t
X
16189
16278
N1a
16147t
16248
N*
16257t
16261
N3
16290
16319
16223
F
A
16304
H
CRS
7028
B
9-bp indel
16189
R*
73
HV*
16298
V
12308
16126
U6
16172
16219
U*
16270
U5
16294
T
16189
16249
16051
16129t
U2
U1
16224
16311
16356
16343
16069
K
U4
U3
J

Table 5.1: Frequency of major mtDNA haplogroup observed in several Southeast Asian populations (Schurr and Wallace, 2002).

Population	n	B	F	M	Total freq
Taiwanese	20	7.1	22.2	35.0	64.3
Malaysian Han	14	32.0	-	35.7	67.7
Vietnamese	28	17.9	32.1	32.1	82.1
Orang Asli	31	3.1	27.3	39.4	69.8
Malays	14	14.3	14.3	57.1	85.7
Sabah Aborigines	32	18.6	2.9	56.3	77.8
Papua New Guinea	119	41.8	1.0	44.5	87.3
Koreans	13	15.4	7.7	61.5	84.6
Tibetans	54	11.1	14.8	61.1	87
Nepalese	50	4.0	4.0	30.0	38

n= number of tested samples

5.2 Restriction Analysis of mtDNA Coding Region

A total of 9 sets of primers (listed in Table 2.1.2) were used for amplification of mtDNA coding region fragments bearing the targeted polymorphic site. These amplified fragments were then subjected to restriction digestion using relevant enzymes in order to determine the polymorphic state. Digested PCR product was resolved by agarose gel electrophoresis and visualised by ethidium bromide staining.

5.2.1 Restriction Analysis of 10394/10397 *Dde* I/*Alu* I Sites (Superhaplogroup M)

The first restriction analysis carried for both Modern Malay and Orang Asli populations involved the RFLP markers used for determination of superhaplogroup M, which is among the most common haplogroup found in Asian population studied so far (Kivisild *et al.*, 1999). Superhaplogroup M is defined by the presence of an *Alu* I site at nucleotide position 10,397, created by a C-to-T transition at nucleotide position 10,400 and a *Dde* I site at nucleotide position 10,394, created by an A-to-G transition at nucleotide position 10,398.

Superhaplogroup M was further divided into several sub-branches characterised by few HVR1 polymorphisms motifs, which slightly differ between populations. Figure 5.2 showed the examples of sub-branches of superhaplogroup M in some Asian populations.

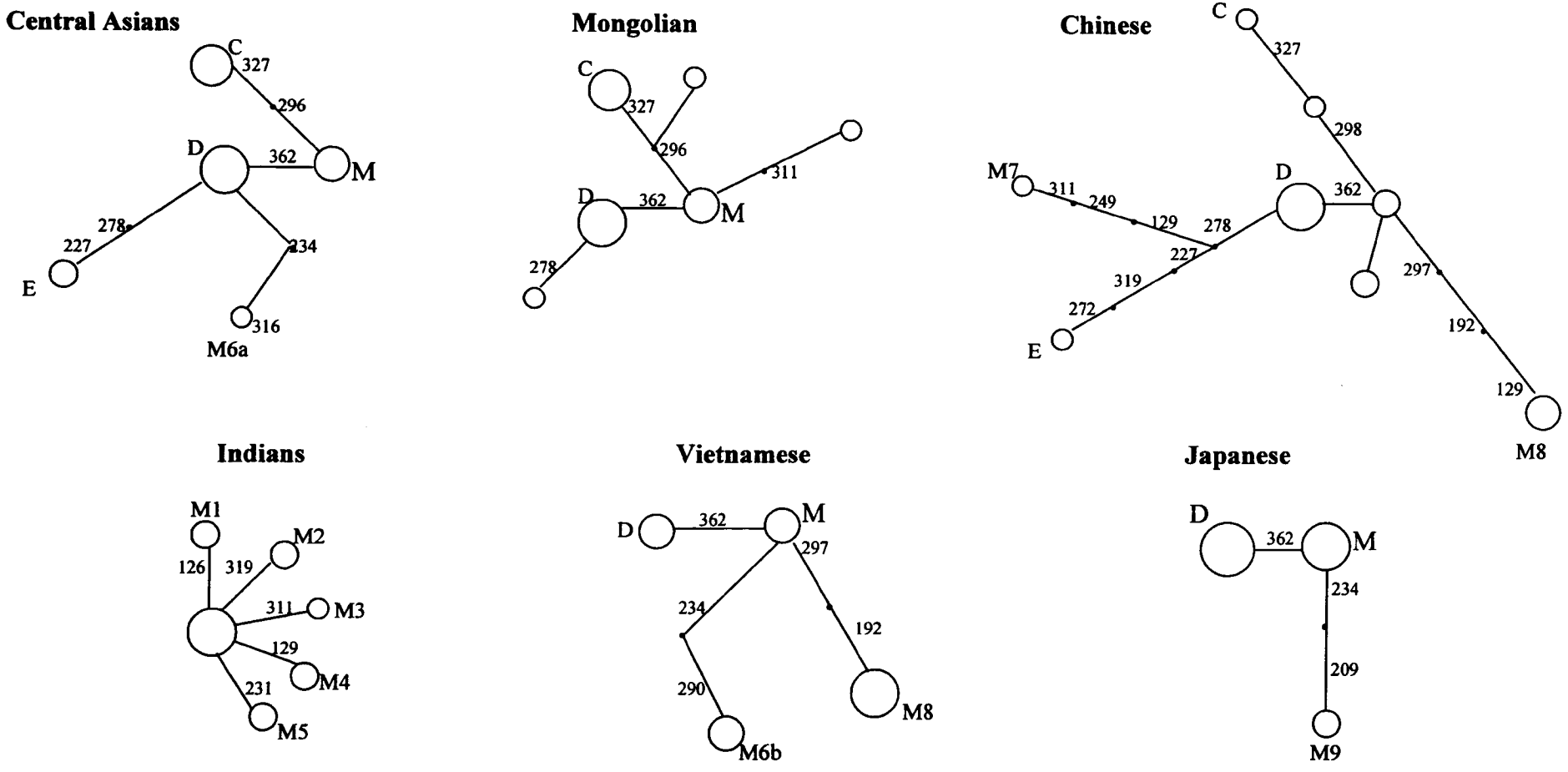
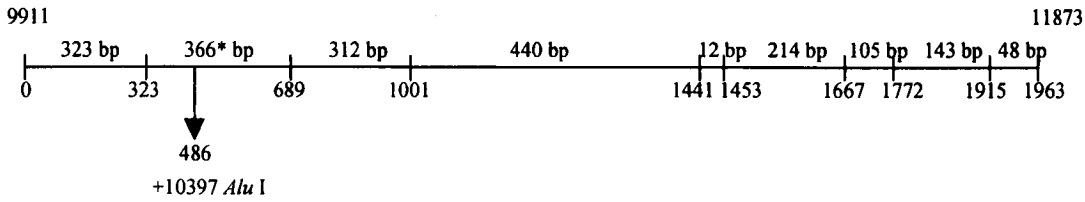


Figure 5.2: Superhaplogroup M sub-structure observed in some Asian populations. HVI polymorphisms used in the classifications are shown along the lines connecting the nodes, with sequence numbers less 16,000. The node area reflects the frequency of each sub-branch (Kivisild *et al.*, 1999).

A fragment of 1.9 kb harbouring both sites was amplified using a set of primers, 9911-F and 11873-R. Gene Jockey software (Taylor, 1990) was used to check for other *Alu* I and *Dde* I restriction sites existed within this amplified fragment by using the Cambridge Reference Sequence (Anderson *et al.*, 1981; Andrews *et al.*, 1999) as a reference for a non-M haplogroup. The cut positions and fragment sizes of the digested products are shown in Figure 5.3 and Figure 5.4. The amplified fragment was then digested using *Alu* I and *Dde* I restriction enzyme in a separate reaction. The presence or absence of this restriction site was determined by the pattern of bands observed after the digested product was resolved by agarose gel electrophoresis (shown in Figure 5.5 and 5.6).

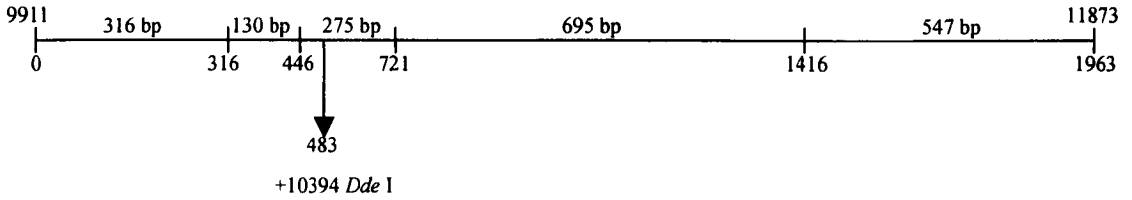
Figure 5.3: Map of *Alu* I digestion for mtDNA coding region fragment amplified using primer 9911-F and 11873-R.



***Alu* I cut position for CRS**	***Alu* I cut position for 10400 C to T**	**Fragment size (bp) for -10397 *Alu* I**	**Fragment size (bp) for +10397 *Alu* I**
323	323	12	12
689	486	48	48
1001	689	105	105
1441	1001	143	143
1453	1441	214	163
1667	1453	312	203
1772	1667	323	214
1915	1772	366*	312
	1915	440	323
			440

If an additional *Alu* I site existed at np 10,397, the 366* bp fragment will be cut into another two fragments of 163 bp and 203 bp. Sample bearing nucleotide transition from C to T at np 10,400, which simultaneously created an additional *Alu* I at np 10,397 can be identified from the lost of this 366* bp fragment when the digested product was resolved using agarose gel electrophoresis.

Figure 5.4: Map of *Dde* I digestion for mtDNA coding region fragment amplified using primer 9911-F and 11873-R.



***Dde* I cut position for CRS**	***Dde* I cut position for 10398 A to G**	**Fragment size (bp) for -10394 *Dde* I**	**Fragment size (bp) for +10394 *Dde* I**
316	316	130	37
446	446	275*	130
721	483	316	238
1416	721	547	316
	1416	695	547
			695

If an additional *Dde* I site existed at np 10,394 (position 483 in the amplified fragment), the 275* bp fragment will be cut into another two fragments of 37 bp and 238 bp. Sample bearing nucleotide transition from A to G at np 10,398, which simultaneously created an additional *Dde* I site at np 10,394 can be differentiated by the lost of this 275* bp fragment when the digested product was resolved by agarose gel electrophoresis.

Figure 5.5: Agarose gel electrophoresis of *Alu* I digestion for mtDNA coding region fragment amplified using primer 9911-F and 11873-R.

Lane 1: 1 kb DNA ladder (1 μg/μl), 1 μl loaded

Lane 2-8: 4 μl of digested product electrophoresed on 1.2% agarose gel at 80 V for 45 min using 1X TBE buffer.

Lane	2	3	4	5	6	7	8
Set 1	+	+	+	-	+	-	+
Set 2	+	+	-	-	-	+	-
Set 3	+	+	-	+	+	-	-
Set 4	-	-	+	+	-	-	-

A plus sign (+) indicates an *Alu* I site gain at np 10,397 while the minus (-) sign indicates an *Alu* I site loss at np 10,397 within the amplified fragment.

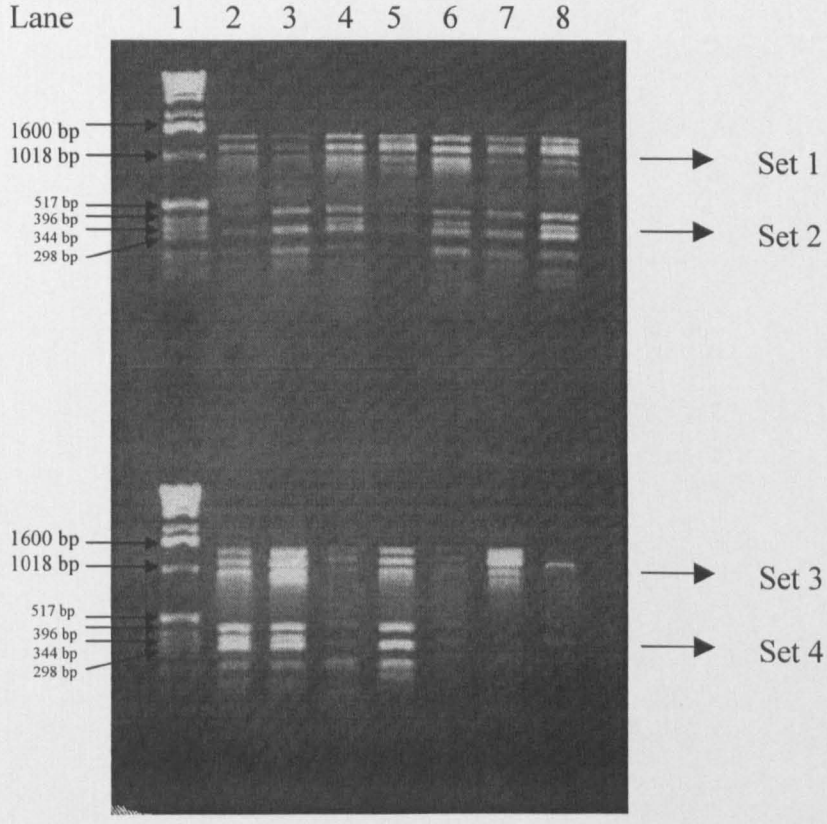

Lane
1 2 3 4 5 6 7 8
1600 bp
1018 bp
517 bp
396 bp
344 bp
298 bp
1600 bp
1018 bp
517 bp
396 bp
344 bp
298 bp
Set 1
Set 2
Set 3
Set 4

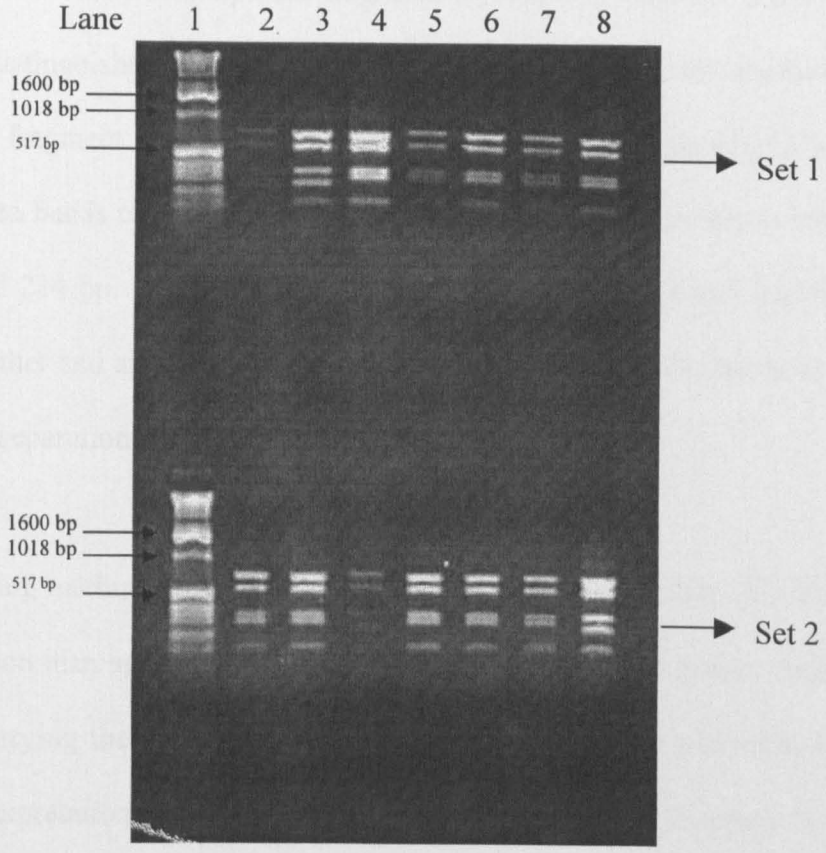
Figure 5.6: Agarose gel electrophoresis showing the digested product of mtDNA coding region fragment amplified using primer 9911-F and 11873-R with *Dde* I restriction enzyme.

Lane 1: 1 kb DNA ladder (1 μg/μl), 1 μl loaded

Lane 2-8: 4 μl of digested product electrophoresed on 1.2% agarose gel at 80 V for 45 min using 1X TBE buffer.

Lane	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Set 1	+	+	-	-	-	-	-
Set 2	-	-	+	-	-	-	+

A plus sign (+) indicates a *Dde* I site gain at np 10,394 while the minus (-) sign indicates a *Dde* I site loss at np 10,394 within the amplified fragment.


Lane 1 2 3 4 5 6 7 8
1600 bp
1018 bp
517 bp
Set 1
1600 bp
1018 bp
517 bp
Set 2

The digestion pattern of fragment amplified using primer 9911-F and 11873-R was not showing noticeable difference between samples with the presence or absence of the 10,397 *Alu* I site and 10,394 *Dde* I site. This was due to the occurrence of additional *Alu* I and *Dde* I sites within the amplified fragment. Samples carrying the 10,397 *Alu* I site can only be distinguished by the lost of a 366 bp fragment from the resolved digested product. This fragment was further cut into two smaller fragments of 163 bp and 203 bp. These extra bands cannot be seen since there were another two bands with the size of 143 bp and 214 bp. The small size difference between these bands had made them migrated together and appeared as single band since agarose gel electrophoresis did not provide good separation for low molecular weight DNA.

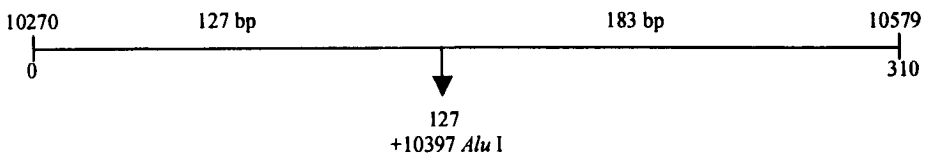
Higher resolving medium such as NuSieve agarose or polyacylamide gel should served better separation than agarose gel. This method was however not tested. Determination of samples carrying the 10,394 *Dde*I site had also faced similar problems. In order to make data interpretation much easier and clearer, another shorter fragment bearing both 10,397 *Alu* I and 10,394 *Dde* I sites was amplified using 10291-F and 10556-R primers.

This new fragment totally eliminates other *Alu* I sites and had only one additional *Dde* I site at np 10,356. Identification for samples bearing the 10,397 *Alu* I site was much easier since the amplified fragment for sample not bearing this polymorphic site had stayed intact after digestion due to the *Alu* I site lost. Samples carrying the 10,394 *Dde* I site was also easily distinguished by the clear difference of the band pattern between different variants.

Restriction digestion map of *Alu* I and *Dde* I digestion for this shorter fragment is shown in Figure 5.7. The resolved digested product is shown in Figure 5.8. From this restriction analysis, 50 Modern Malay and 24 Orang Asli samples were identified to carry both 10,394/10,397 *Dde* I/*Alu* I restriction sites. Fifteen of the Modern Malay and 32 Orang Asli were revealed to bear only the 10,394 *Dde* I site. All samples showing the presence of 10,394/10,397 *Dde*I /*Alu*I sites were classified into superhaplogroup M.

Figure 5.7: Map of *Alu* I and *Dde* I digestion for mtDNA coding region fragment amplified using primer 10291-F and 10556-R.

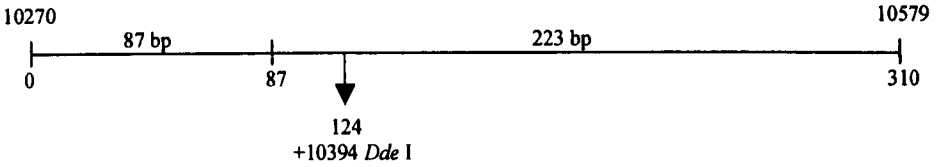
***Alu* I digestion:**



*Alu*I cut position for CRS	*Alu*I cut position for 10400 C to T	Fragment size (bp) for -10397 *Alu* I	Fragment size (bp) for +10397 *Alu* I
-	127	310	127 183

If an *Alu* I site existed at np position 10,397, the 310 bp fragment will be cut into two smaller fragments of 127 bp and 183 bp. Sample without this polymorphism will stay as an intact band after the digestion due to the *Alu* I site lost.

***Dde* I digestion:**



Dde I cut position for CRS	*Dde* I cut position for 10398 A to G	Fragment size (bp) for -10394 *Dde* I	Fragment size (bp) for +10394 *Dde* I
87	87 124	87 223*	37 87 186

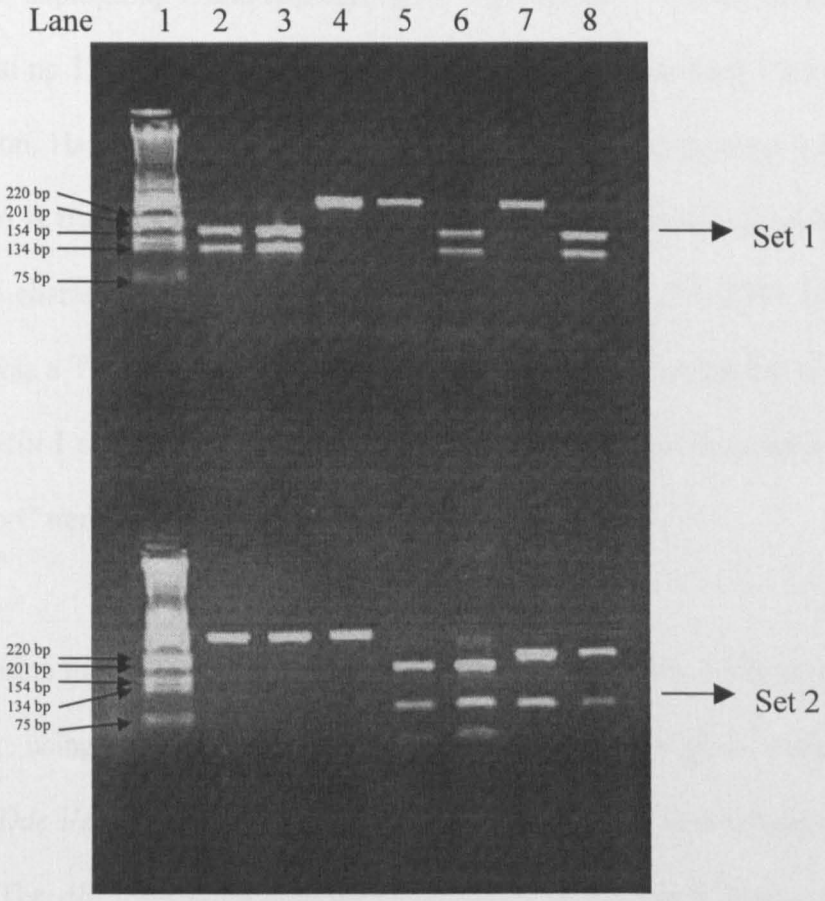
An additional *Dde* I site existing at np 10,394 will cut this 310 bp fragment into 3 smaller fragments of 37 bp, 87 bp and 186 bp and simultaneously eliminate the 223* bp fragment. Samples carrying this polymorphism are easily distinguished by the appearance of these three bands in agarose gel electrophoresis.

Figure 5.8: Agarose gel electrophoresis showing the digested product of *Alu* I and *Dde* I digestion for mtDNA coding region fragment amplified using primer 10291-F and 10556-R.

Lane 1: 1 kb DNA ladder (1 μg/μl), 1 μl loaded

Lane 2-8: 4 μl of digested product electrophoresed on 1.2% agarose gel (ethidium bromide stained) at 80 V for 45 min using 1X TBE buffer.

Set	**1**	**2**
Lane 2	+10397 *Alu* I	-10397 *Alu* I
Lane 3	+10397 *Alu* I	-10397 *Alu* I
Lane 4	-10397 *Alu* I	-10397 *Alu* I
Lane 5	-10397 *Alu* I	+10394 *Dde* I
Lane 6	+10397 *Alu* I	+10394 *Dde* I
Lane 7	-10397 *Alu* I	-10394 *Dde* I
Lane 8	+10397 *Alu* I	-10394 *Dde* I

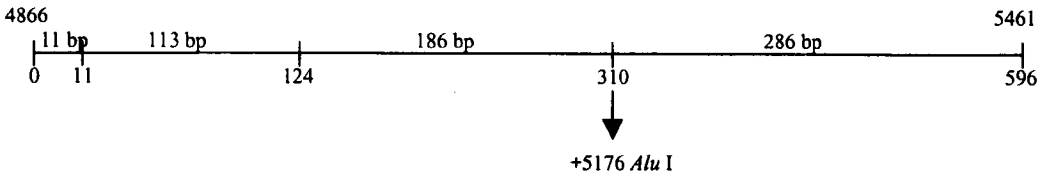

Lane 1 2 3 4 5 6 7 8
220 bp
201 bp
154 bp
134 bp
75 bp
Set 1
220 bp
201 bp
154 bp
134 bp
75 bp
Set 2

5.2.2 Restriction Analysis of 5,176 and 13,262 *Alu* I Sites (Haplogroup D and C)

All samples that previously classified into superhaplogroup M were further analysed for determination of haplogroup C and haplogroup D. Haplogroup C was determined by an *Alu* I site gain at np 13,262, which created by an A to G transition at np 13,263 within the coding region. Haplogroup D was classified by an *Alu* I site loss at np 5,176, as a result from C to A transversion at np 5,178. Besides this RFLP marker, haplogroup C and D was also characterised by HVI sequence motifs. The definitive HVI marker for haplogroup C was a T to C transition at np 16,298 and a C to T transition at np 16,327. In addition to *Alu* I site loss at np 5,176, haplogroup D was also determined by the presence of T to C transition at np 16,362 within the HVI region.

In order to examine the presence or absence of *Alu* I site at np 5,176, a 596 bp fragment was amplified using 4887-F and 5442-R primers. All samples carrying the 10,394/10,397 *Dde* I/*Alu* I sites and 16,362 nucleotide transition were screened for this RFLP marker. The *Alu* I digestion map for this amplified fragment is shown in Figure 5.9. Electrophoresis of the digested product is shown in Figure 5.10. From the resolved digested product, sample carrying C to A transversion at np 5,178 was easily identified by the presence of a 472 bp fragment while the other variant had lost this fragment due to the existence of additional 5,176 *Alu* I site. The only problem observed was the inconsistency of the digestion pattern for sample bearing the 5,176 *Alu* I site. This was due to multiple variant of *Alu* I cut position within the amplified fragment. To resolve this ambiguity, a shorter fragment was amplified using a new forward primer (5160-F), which manages to eliminate all insignificant *Alu* I sites. The new *Alu* I restriction map for this 322 bp fragment is shown in Figure 5.11.

Figure 5.9: Map of *Alu* I digestion for mtDNA coding region fragment amplified using primer 4887-F and 5442-R.



Alu I cut position for CRS	*Alu* I cut position for 5178 C to A	Fragment size (bp) for +5176 *Alu* I	Fragment size (bp) for -5176 *Alu* I
11	11	11	11
124	124	113	113
310		186	472*
		286	

Nucleotide transversion from C to A at np 5178 had concurrently led to an *Alu* I site loss at np 5,176. Digested product resolved by agarose gel electrophoresis for sample carrying this polymorphism had only shown 2 bands of 113 bp and 472 bp. The 11 bp band could not be seen due to its small size. Sample not bearing this polymorphism had an additional *Alu* I site at np 5,176 and therefore lost the 472* bp fragment.

Figure 5.10: Agarose gel electrophoresis showing the digested product for the mtDNA coding region fragment amplified using primer 4887-F and 5442R.

Lane 1: 1 kb DNA ladder (1 μg/μl), 1 μl loaded

Lane 2-8: 4 μl of digested product electrophoresed on 1.2% agarose gel (ethidium bromide stained) at 80 V for 45 min using 1X TBE buffer.

Lane	2	3	4	5	6	7	8
Set 1	+	+	+	+	+	+	+
Set 2	+	+	+	+	+	+	-

The plus sign (+) indicates the existence of an *Alu* I site at np 5,176 while the minus sign (-) indicates an *Alu* I site loss at np 5,176 within the amplified fragment. Note the difference between the digestion pattern for sample in lane 3 (for both sets) and lane 7 to other samples bearing the 5,176 *Alu* I site.

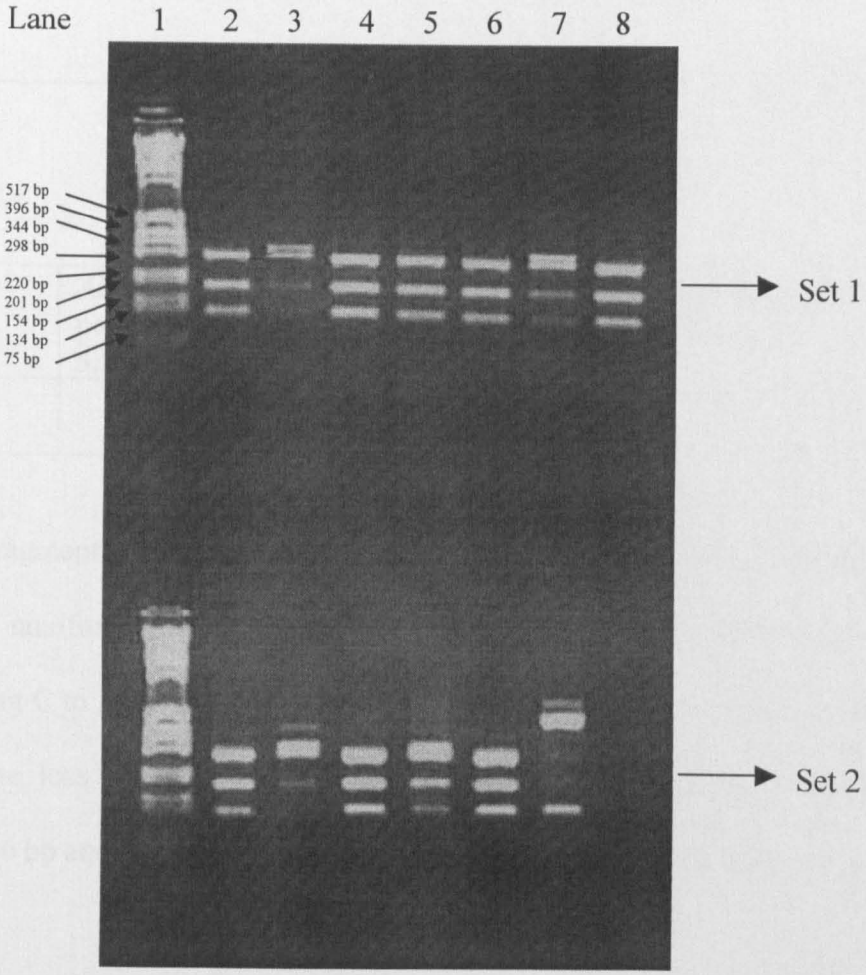
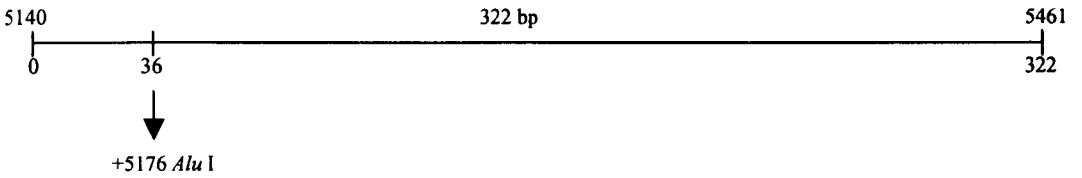

Lane 1 2 3 4 5 6 7 8
517 bp
396 bp
344 bp
298 bp
220 bp
201 bp
154 bp
134 bp
75 bp
Set 1
Set 2

Figure 5.11: Map of *Alu* I digestion for mtDNA coding region fragment amplified using primer 5160-F and 5442-R.



***Alu* I cut position for CRS**	***Alu* I cut position for 5,178 C to A**	**Fragment size (bp) for +5176 *Alu* I**	**Fragment size (bp) for -5176 *Alu* I**
36	-	36 286	322

This shorter fragment (322 bp) provided a straightforward identification of haplogroup D since other uninformative *Alu* I sites were eliminated. The amplified fragment for samples bearing C to A transversion at position 5,178 had stay intact after the digestion due to the site loss. Samples not belong to haplogroup D contained two smaller fragments of 36 bp and 286 bp, as a result from the presence of 5,176 *Alu* I site.

All samples carrying 10,394/10,397 *Dde* I/*Alu* I sites and 16,362 HVI motif were screened again for 5,176 *Alu* I site by using the shorter amplified fragment, which gave a straightforward interpretation for haplogroup D classification. The resolved digested product of this amplified fragment is shown in figure 5.12. Based on this restriction digestion analysis, two samples of the Modern Malay population were classified into haplogroup D. Both samples (M83 and M92) exhibit the 16,362 T to C nucleotide transition, a definitive HVI motif for haplogroup D. None of the Orang Asli sample was found to belong to this haplogroup.

A fragment of 835 bp was amplified using 12978-F and 13773-R primers for determination of haplogroup C. Restriction digestion map for this amplified fragment is shown in Figure 5.13. Samples carrying 13,263 A to G transition, which simultaneously created an *Alu* I site at np 13,262 was determined by the appearance of 304 bp and 531 bp fragments from the resolved digested product. Sample with the absence of this restriction site had stayed intact after the digestion. Figure 5.14 showing the agarose gel electrophoresis of *Alu* I digestion for the amplified fragment. Only one sample from the Modern Malay population was classified into haplogroup C. This sample had also contained both definitive HVI markers used to define haplogroup C. None of the Orang Asli samples were classified as haplogroup C.

Figure 5.12: Agarose gel electrophoresis showing the resolved digested product of *Alu* I digestion for mtDNA coding region fragment amplified using primer 5160-F and 5442-R.

Lane 1: 1 kb DNA ladder (1 μg/μl), 1 μl loaded.

Lane 2-8: 4 μl of digested product electrophoresed on 1.2% agarose gel (ethidium bromide stained) at 80 V for 45 min using 1X TBE buffer.

Lane	2	3	4	5	6	7	8
Set 1	+	-	+	+	+	+	+
Set 2	+	+	+	+	+	+	-

The plus sign (+) indicates the existence of an *Alu* I site at np 5,176 within the amplified fragment while the minus sign (-) indicates an *Alu* I site loss at np 5,176, as a result from C to A transversion at np 5,178.

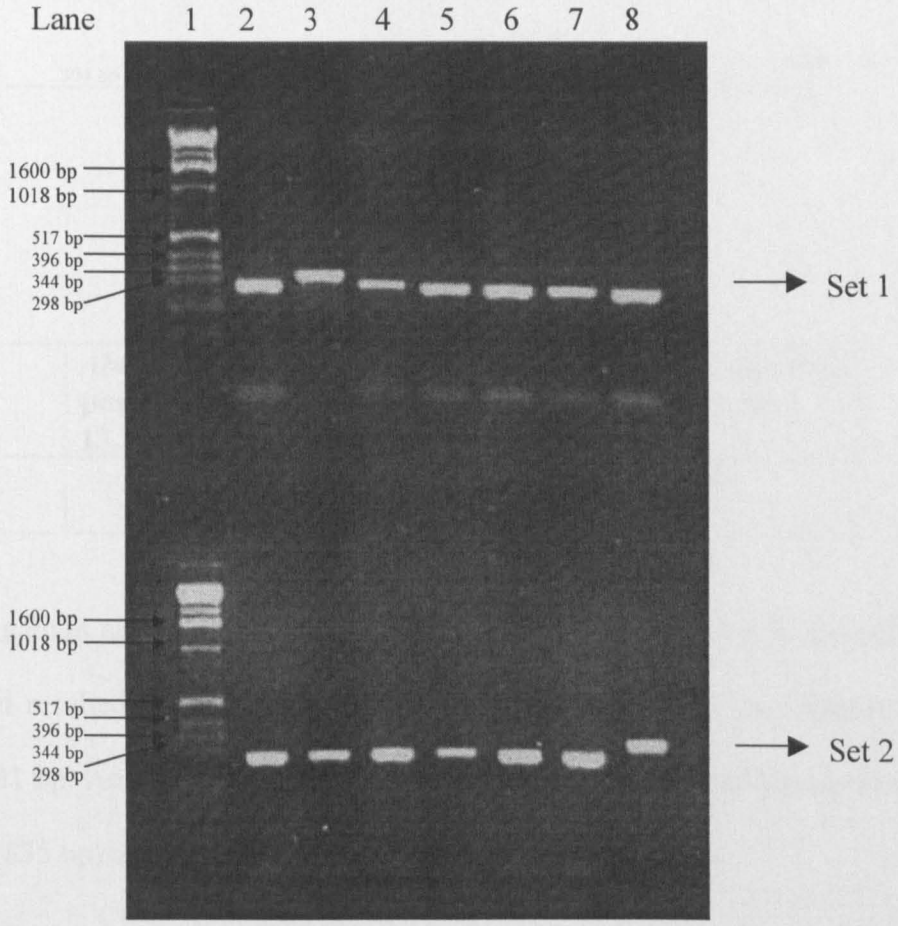
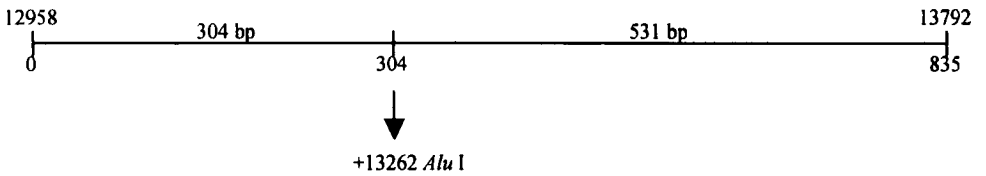

Lane 1 2 3 4 5 6 7 8
1600 bp
1018 bp
517 bp
396 bp
344 bp
298 bp
Set 1
1600 bp
1018 bp
517 bp
396 bp
344 bp
298 bp
Set 2

Figure 5.13: Map of *Alu* I digestion for mtDNA coding region fragment amplified using primer 12978-F and 13773-R.



***Alu* I cut position for CRS**	***Alu* I cut position for 13,263 A to G**	**Fragment size (bp) for +13262 *Alu* I**	**Fragment size (bp) for -13262 *Alu* I**
-	304	304 531	835

An *Alu* I site gain at nucleotide position 13,262 for sample bearing A to G transition at np 13,263 had resulted for digestion of this amplified fragment into two fragments of 304 bp and 531 bp. Amplified fragment for samples not bearing this polymorphism had stayed intact (835 bp) after the digestion.

Figure 5.14: Agarose gel electrophoresis showing the resolved digested product for fragment amplified using primer 12978-F and 13773-R.

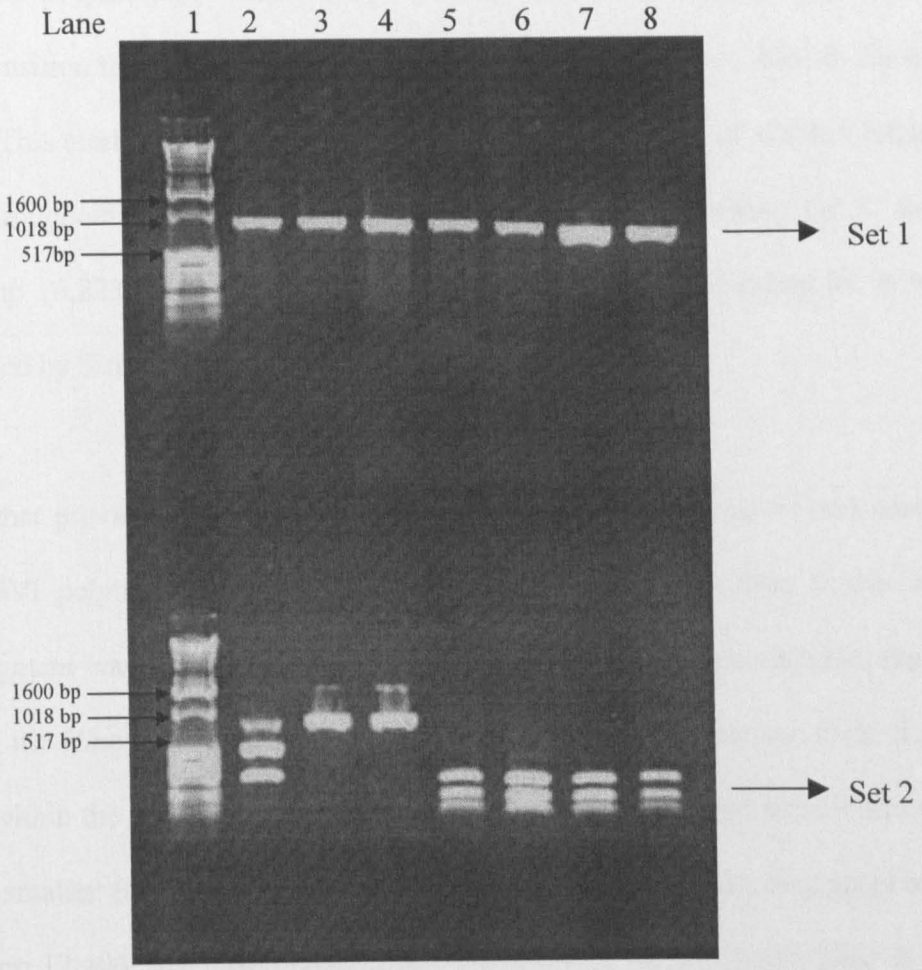
Lane 1: 1 kb DNA ladder (1 μg/μl), 1 μl loaded.

Lane 2-8: 4 μl of digested product electrophoresed on 1.2% agarose gel (ethidium bromide stained) at 80 V for 45 min using 1X TBE buffer.

Lane	2	3	4	5	6	7	8
Set 1	-	-	-	-	-	-	-
Set 2	+	-	-	x	x	x	x

The *Alu* I site gain at np 13,262 was only detected in one sample (shown with + sign in above table) within all screened samples. The minus sign (-) indicates that this fragment was not cut by the *Alu* I restriction enzyme due to the site loss.

x: digested product for different amplified fragment.

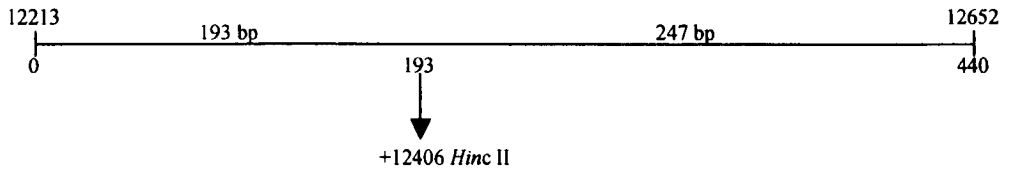

Lane 1 2 3 4 5 6 7 8
1600 bp
1018 bp
517bp
Set 1
1600 bp
1018 bp
517 bp
Set 2

5.2.3 Restriction Analysis of 12,406 *Hinc* II Site (Haplogroup F)

A fragment of 440 bp was amplified using primer 12233-F and 12630-R in order to investigate the polymorphic state at np 12,406 within the sample populations. Nucleotide transition from G to A at this particular position has led to a *Hinc* II site loss at np 12,406. This marker had been used to define haplogroup F, one of mtDNA lineage that found throughout Southeast Asia. Haplogroup F was also lacking the C to T transition at np 16,223, which was a fixed HVI motif for superhaplogroup M. It was instead replaced by T to C transition at np 16,304 in haplogroup F.

All samples that previously determined not belong to superhaplogroup M and carried the 16,304 HVI polymorphism, were further screened for 12,406 *Hinc* II site. The amplified fragment was subjected to digestion using *Hinc* II restriction enzyme. Figure 5.15 showing the *Hinc* II digestion map for this 440 bp fragment. Only one *Hinc* II site was located within the amplified fragment. Sample bearing the 12,406 *Hinc* II site was cut into two smaller fragments of 193 bp and 247 bp, while sample carrying G to A transition at np 12,406 had stayed intact after digestion due to the 12,406 *Hinc* II site loss. Figure 5.16 shows the digestion pattern of the resolved digested product. From 24 Modern Malay samples screened, 17 were classified into haplogroup F. The remaining samples were revealed to carry 12,406 *Hinc* II site. For the Orang Asli population, all 28 samples screened contained the 12,406 *Hinc* II site and therefore did not belong to haplogroup F.

Figure 5.15: Map of *Hinc* II digestion for mtDNA coding region fragment amplified using primer 12233-F and 12630-R.



***Hinc* II cut position for CRS**	***Hinc* II cut position for 12,406 G to A**	**Fragment size (bp) for +12406 *Hinc* II**	**Fragment size (bp) for -12406 *Hinc* II**
193	-	193 247	440

The amplified fragment for samples carrying G to A transition at np 12,406 stayed intact (440 bp) after *Hinc* II digestion due to the site loss. Amplified fragments for sample not bearing this polymorphism was cut into two smaller fragments of 193 bp and 247 bp. Both variants were clearly distinguished by agarose gel electrophoresis.

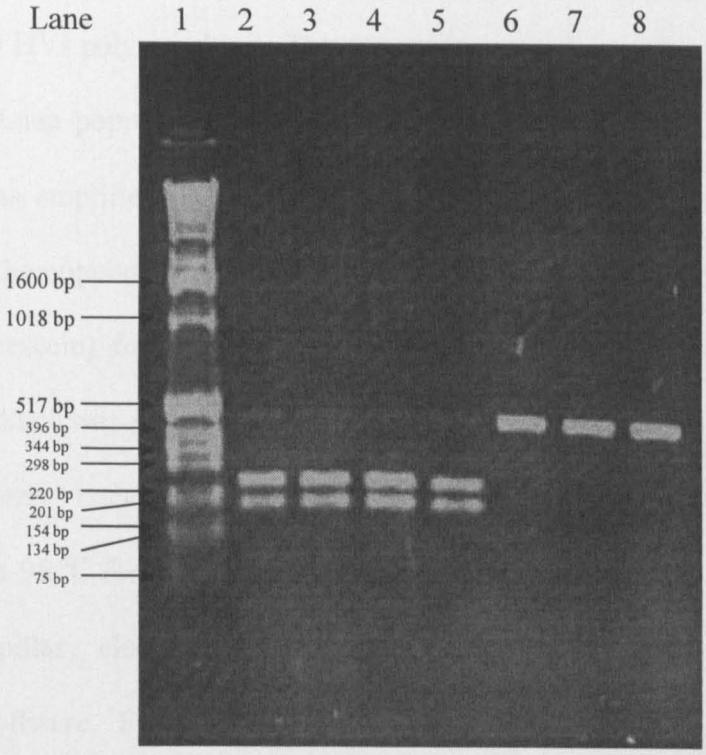
Figure 5.16: Agarose gel electrophoresis showing the resolved digested product for fragment amplified using primer 12233-F and 12630-R.

Lane 1: 1 kb DNA ladder (1 μg/μl), 1 μl loaded.

Lane 2-8: 4 μl of digested product electrophoresed on 1.2% agarose gel (ethidium bromide stained) at 80 V for 45 min using 1X TBE buffer.

Lane	2	3	4	5	6	7	8
Digestion status	+	+	+	+	-	-	-

The plus sign (+) indicates that the sample was digested by *Hinc* II restriction enzyme (+12406 *Hinc* II) while the minus sign (-) indicates the site loss within the amplified fragment (-12406 *Hinc* II).

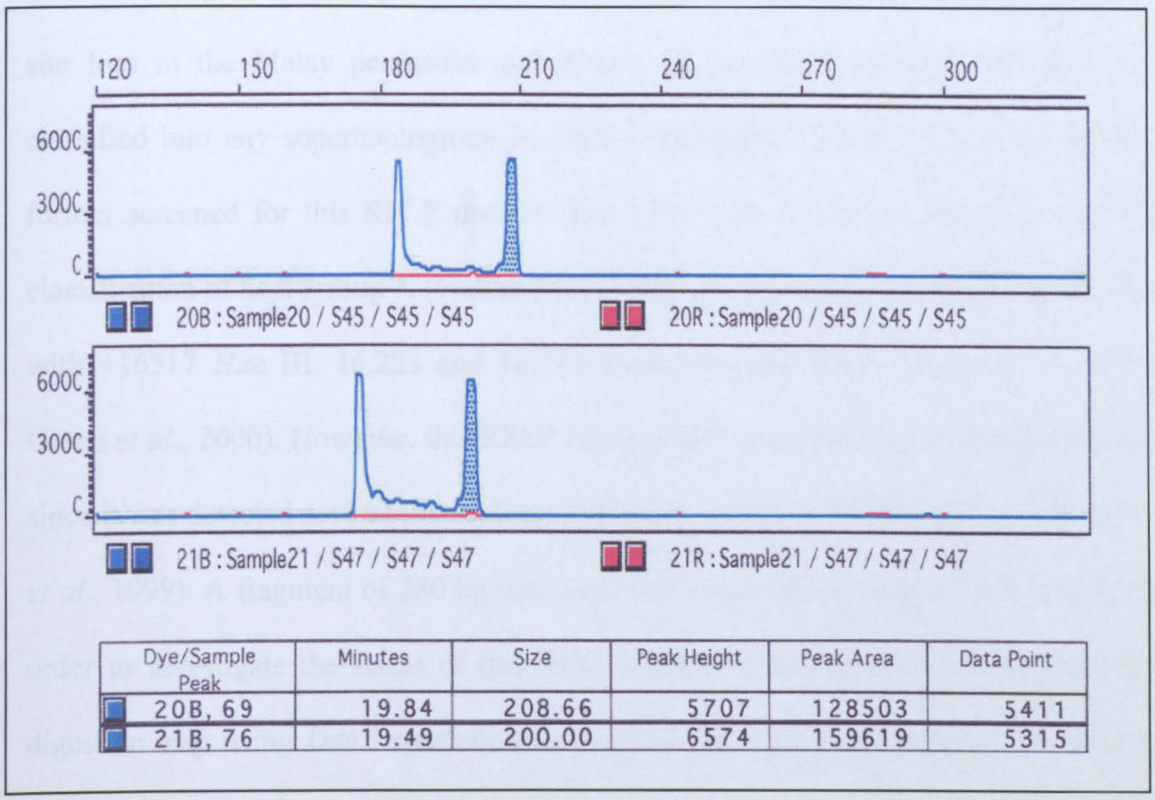

Lane
1
2
3
4
5
6
7
8
1600 bp
1018 bp
517 bp
396 bp
344 bp
298 bp
220 bp
201 bp
154 bp
134 bp
75 bp

5.2.4 9-bp COII-tRNA^{Lys} Deletion (Haplogroup B)

Haplogroup B is distinguished by the 9-bp deletion occurring between np 8,272-8,289 in the COII/tRNA^{Lys} region (Torroni *et al.*, 1992). It was always found in association with 16,189 HVI polymorphism. This 9-bp deletion is a common marker established in Southeast Asian populations (Melton *et al.*, 1995). A fragment of 211 bp (including primers) was amplified using 8215-F and 8385-R primers for screening of the 9-bp deletion in the population samples. The forward primer was fluorescently labelled using FAM (fluorescein) for size detection in the ABI PRISM® 310 Genetic Analyzer. A loading cocktail was prepared by adding 0.5 µl PCR product, 0.5 µl Genescan™ 500 (ROX) Internal Lane Standard and 9 µl deionised formamide. This mixture was denatured at 95 °C for 3 min and snap-cooled on ice for another 3 min before loading into the capillary electrophoresis system. Samples were analysed using GeneScan® Analysis software. Figure 5.17 shows the electropherogram of samples with the presence and absence of the 9-bp deletion.

All samples that were previously not classified into superhaplogroup M clusters and haplogroup F were screened for the 9-bp deletion. 19 samples from the Modern Malay population and 4 samples from the Orang Asli population were identified with the deletion. All samples detected with the 9-bp deletion at the same time carried the 16,189 T to C transition. Based on the occurrence of the 9-bp deletion along with the definitive HVI polymorphism, these samples were classified into haplogroup B.

Figure 5.17: Electropherogram showing peaks of the labelled amplified fragment for the 9-bp deletion detection.



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
20B, 69	19.84	208.66	5707	128503	5411
21B, 76	19.49	200.00	6574	159619	5315

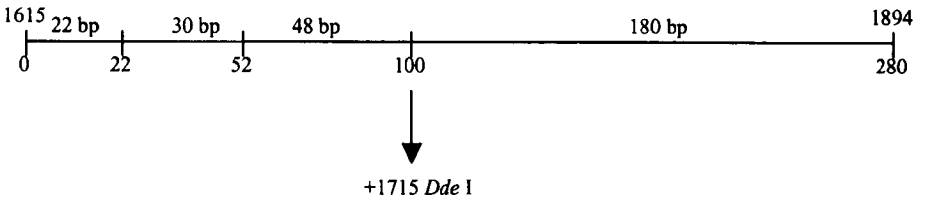
The top electropherogram represent sample without the 9-bp deletion (M45) while the bottom electropherogram showing the profile for sample with the 9-bp deletion (M47). Note that the difference between the labelled peak sizes was approximately 9 bp (8.66-bp in the size calling data), indicating the existence of the 9-bp deletion in sample M47. A second peak (not labelled in the electropherogram), was observed in all screened samples. This extra peak might be a result from excessive amount of DNA used in the PCR reaction. The appearance of this peak was eventually reduced when less DNA template was used. Since the existence of this peak did not complicate data interpretation, this result was therefore accepted.

5.2.5 Restriction Analysis of 1715 *Dde* I Site

Based on a report by Ballinger *et al.*, 1992, which indicated the presence of 1,715 *Dde* I site loss in the Malay peninsular population, all remaining samples that were not classified into any superhaplogroup M cluster, haplogroup F and haplogroup B were further screened for this RFLP marker. The 1,715 *Dde* I site loss was also used for classification of haplogroup X (Forster *et al.*, 1996; Torroni *et al.*, 1996), in concurrence with +16517 *Hae* III, 16,223 and 16,278 nucleotide transition (Brown *et al.*, 1998; Green *et al.*, 2000). However, this RFLP site was also considered as recurrent mutation since it was detected several times along with other definitive RFLP markers (Macaulay *et al.*, 1999). A fragment of 280 bp was amplified using 1643-F and 1874-R primers in order to investigate the status of this RFLP marker in the unclassified samples. The digestion map using *Dde* I restriction enzyme for this amplified fragment is shown in Figure 5.18. A total of 15 Modern Malay samples and 30 Orang Asli samples were screened for the 1,715 *Dde* I site. The digestion pattern of the resolved digested products are shown in Figure 5.19.

Restriction analysis of 1,715 *Dde* I site had successfully identified 2 samples in the Modern Malay population and 28 samples in the Orang Asli population with the site loss. All samples detected with the site loss did not carry any of the HVI motifs for haplogroup X and therefore did not designated into this haplogroup. These samples were instead showing the combination of 16,295, 16,296 and 16,304 nucleotide transition in concurrence with the 1,715 *Dde* I site loss. Based on the HVI sequence motifs, these samples were classified into haplogroup R21 (Martin Richards, unpublished data).

Figure 5.18: Map of *Dde* I digestion for mtDNA coding region fragment amplified using primer 1643-F and 1874-R.



***Dde* I cut position for CRS**	***Dde* I cut position for 1,719 G to A**	**Fragment size (bp) for +1715 *Dde* I**	**Fragment size (bp) for -1715 *Dde* I**
22	22	22	22
52	52	30	30
100		48	228
		180	

Nucleotide transition from G to A at np 1,719 led to *Dde* I site loss at np 1,715. Amplified fragment carrying this transition was cut into three fragments. However, only two bands were observed by agarose gel electrophoresis since the 22 bp and 30 bp bands appeared as single band due to the small size difference.

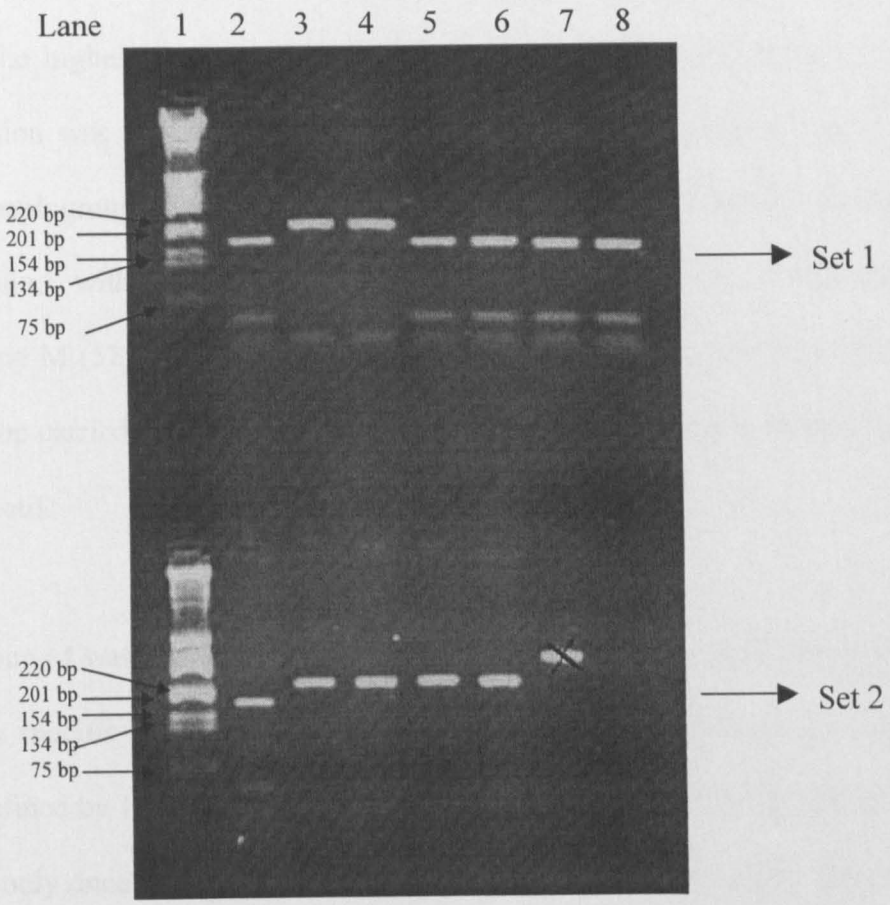
Figure 5.19: Agarose gel electrophoresis showing the resolved digested product of *Dde* I digestion for fragment amplified using primer 1643-F and 1874-R.

Lane 1: 1 kb DNA ladder (1 μg/μl), 1 μl loaded.

Lane 2-8: 4 μl of digested product electrophoresed on 1.2% agarose gel (ethidium bromide stained) at 80 V for 45 min using 1X TBE buffer.

Lane	2	3	4	5	6	7	8
Set 1	+	-	-	+	+	+	+
Set 2	+	-	-	-	-		

The plus sign (+) indicates the presence of *Dde* I site at np 1,715 (+1715 *Dde* I) while the minus sign (-) indicates the site loss within the amplified fragment (-1715 *Dde* I).


Lane 1 2 3 4 5 6 7 8
220 bp
201 bp
154 bp
134 bp
75 bp
Set 1
220 bp
201 bp
154 bp
134 bp
75 bp
Set 2

5.3 Discussion

RFLP analysis of mtDNA coding region for the Modern Malay and Orang Asli population has successfully classified most of the samples into a specific mtDNA haplogroup. The highest frequency of mtDNA haplogroup observed in the Modern Malay population was superhaplogroup M, with 49% of the samples belong to this cluster. Superhaplogroup M was also observed at high frequency (41.38%) in the Orang Asli population, with the Kinsiu sub-group showing higher frequency of superhaplogroup M (57.14%) than the Jahai (26.67%). All samples classified into this superhaplogroup carried both 10,394/10,397 *Dde* I/*Alu* I site gain and exhibited a fixed 16,223 HVI motif.

Superhaplogroup M was further divided into discrete sub-clusters C, D, E and G, which are defined by specific RFLP markers and HVI sequence polymorphisms. Sub-cluster C, which is defined by 13,262 *Alu* I site gain along with 16,298 and 16,327 HVI motifs, was observed only once within the Modern Malay population. None of this sub-cluster was observed in the Orang Asli population. Sub-cluster D, which is defined by an *Alu* I site loss at np 5,176 and 16,362 HVI motif, occurred twice in the Modern Malay population and again none was detected in the Orang Asli population. Seven of the Modern Malay samples were assigned into sub-cluster E1 by the HVI motif observed (16,390). The *Hha* I site loss at np 7,598, which is used for haplogroup E determination was however not tested for these samples. All samples bearing the 16,390 polymorphism were instead examined in the SNaPshot analysis to confirm the occurrence of G to A transition at np 7,598, which responsible for the *Hha* I site loss.

Three Indian sub-clusters M2, M3 and M4 (Kivisild *et al.*, 1999), which are defined by HVI polymorphism motifs, were also found in the Modern Malay samples. Sub-cluster M2, defined by 16,319 T to C transition, was reported 4 times in the Modern Malay population. Sub-cluster M3 and M4, defined by 16,311 T to C and 16,129 G to A transition, respectively, were accounted 6 times in the Modern Malay samples. Only 2 of the Orang Asli samples were assigned into sub-cluster M4. None of the Orang Asli samples was found to belong to sub-cluster M2 and M3.

The largest superhaplogroup M sub-cluster found within the Orang Asli population was M21a, defined by two HVI polymorphisms at np 16,256 and 16,271 (Martin Richards, unpublished data). This new defined sub-cluster was found to be common in most Semang sub-groups and was also present in the Semang of southern Thailand (Fucharoen *et al.*, 2001). The Kinsiu sub-group showed higher frequency of M21a (46.43%) than the Jahai (23.33%). Only six of the Modern Malay samples were located into sub-cluster M21a, suggesting that it may be an indigenous Semang haplogroup. The final superhaplogroup M sub-cluster observed in the sample population was M7c1c, defined by 16,295 HVI motif (Kivisild *et al.*, 2002; Yao *et al.*, 2002). Three of the Modern Malay samples were positioned into this sub-cluster. None of the Orang Asli samples was found to belong to this sub-cluster.

The Modern Malays were revealed to have a fairly high number of haplogroup F (16.67% of the population), which is a common haplogroup in the Chinese populations and accounted at the highest frequency in Vietnamese population (Horai *et al.*, 1996; Nishimaki *et al.*, 1999; Ballinger *et al.*, 1992). Haplogroup F was defined by a *Hinc* II site loss at np 12,406, along with 16,304 HVI motif and further divided into several sub-

clusters by additional HVI polymorphisms. Sub-cluster F1a, F1a1 and F1a1a were defined by 16,129, 16,162 and 16,108 HVI motifs, respectively (Martin Richards, unpublished data). Sub-cluster F1a was found at the frequency of 4.9%, F1a1 was found at 3.92% and F1a1a was found at 7.84% in the Modern Malay population. None of the Orang Asli samples was classified into haplogroup F.

The 'Asian-specific' 9-bp deletion occurring in the COII/tRNA^{Lys} region of the mtDNA genome, a common marker for the Southeast Asian populations (Melton *et al.*, 1995) was found at the rate of 18.63% in the Modern Malay population. This specific deletion determines haplogroup B, which also occurred in the population of an Asian origin, including Polynesians (Hertzberg *et al.*, 1989; Hagelberg and Clegg, 1993; Lum *et al.*, 1994; Redd *et al.*, 1995) and Native Americans (Schurr *et al.*, 1990; Shields *et al.*, 1992; Torroni *et al.*, 1992; Horai *et al.*, 1993). Haplogroup B was further divided into several sub-clusters by HVI motifs. Sub-cluster B4, B4a and B4b1 was defined by 16,217, 16,261 and 16,136 HVI polymorphism, respectively (Martin Richards, unpublished data). Sub-cluster B4 was found at the frequency of 5.88% while sub-cluster B4a and B4b1 was found at the rate of 0.98% each in the Modern Malay population. Sub-cluster B5a and B5b, which defined by 16,266A and 16,243 HVI motifs, was found at the rate of 9.8% and 0.98%, respectively in the Modern Malays. Only 6.9% of the Orang Asli samples were belong to sub-cluster B5a.

A new haplogroup, R21, was defined by the HVI polymorphisms at np 16,295, 16,296 and 16,304 (Martin Richards, unpublished data). The *Dde* I site loss at np 1,715 was also detected in all samples bearing these polymorphisms. R21 haplogroup appeared very frequently in both Orang Asli sub-groups and emerged only twice in the Modern

Malays. Haplogroup R21 reached the highest frequency in Jahai sub-group, with 66.67% of the samples belongs to this haplogroup. A lower frequency of haplogroup R21 is observed in the Kinsiu sub-group (28.57%) compared to the Jahai.

By referring to the HVI polymorphisms, 2 of the Modern Malay samples were classified into haplogroup Y. HVI motifs used for classification of haplogroup Y involved nucleotide transition from T to C at np 16,126 and 16,231 (Yao *et al.*, 2002). Another 3 samples from the Modern Malay population and 2 samples of the Orang Asli were classified into haplogroup N9a since they exhibited 16,257A and 16,261 HVI polymorphisms that being used for determination of this haplogroup (Yao *et al.*, 2002). R1 haplogroup, determined by 16,266 HVI motif (Kivisild *et al.*, 1999), was observed only once in the Modern Malay population. The remaining 7 samples of the Modern Malay population, which cannot be assigned into any haplogroup, either by RFLP polymorphisms or HVI motifs, were temporarily classify as haplogroup R. Frequency of each mtDNA haplogroup observed within the Modern Malay and Orang Asli population was summarised in Table 5.2. The RFLP polymorphisms, along with HVI sequence variance and haplogroup assigned for each Modern Malay and Orang Asli samples is shown in Table 5.3 and Table 5.4.

Table 5.2: mtDNA haplogroup frequency observed in the Modern Malay and Orang Asli populations of Peninsular Malaysia.

Haplogroup	RFLP polymorphisms	Population frequencies (%)	
		Malay	Orang Asli
M	+10394/+10397 *Dde* I/ *Alu* I	14.71	1.72
M21a	+10394/+10397 *Dde* I/ *Alu* I	5.88	34.48
M21b	+10394/+10397 *Dde* I/ *Alu* I	-	1.72
M7c1c	+10394/+10397 *Dde* I/ *Alu* I	2.94	-
M2	+10394/+10397 *Dde* I/ *Alu* I	3.92	-
M3	+10394/+10397 *Dde* I/ *Alu* I	5.88	-
M4	+10394/+10397 *Dde* I/ *Alu* I	5.88	3.45
M-D	+10394/+10397 *Dde* I/ *Alu* I, -5176 *Alu* I	1.96	-
M-C	+10394/+10397 *Dde* I/ *Alu* I, +13262 *Alu* I	0.98	-
M-E	+10394/+10397 *Dde* I/ *Alu* I, -7598 *Alu* I	6.86	-
B4	9-bp deletion	5.88	-
B4a	9-bp deletion	0.98	
B4b1	9-bp deletion	0.98	
B5a	9-bp deletion	9.80	6.90
B5b	9-bp deletion	0.98	-
F1a	-10394/-10397 *Dde* I/ *Alu* I, -12406 *Hinc* II	4.90	-
F1a1	-10394/-10397 *Dde* I/ *Alu* I, -12406 *Hinc* II	3.92	-
F1a1a	-10394/-10397 *Dde* I/ *Alu* I, -12406 *Hinc* II	7.84	-
N9a	-10394/-10397 *Dde* I/ *Alu* I	2.94	3.45
R	-10394/10397 *Dde* I/ *Alu* I	7.84	-
R1	-10394/-10397 *Dde* I/ *Alu* I	0.98	
R21	-10394/-10397 *Dde* I/ *Alu* I, -1715 *Dde*I	1.96	48.28
Y	-10394/-10397 *Dde* I/ *Alu* I	1.96	-

Table 5.3: RFLP polymorphisms, HVI sequence variance and mtDNA haplogroups of each Modern Malay samples.

SAMPLE	HAPLOTYPES	10397 *Alu* I	10394 *Dde* I	5176 *Alu* I	13262 *Alu* I	1715 *Dde* I	12406 *Hinc* II	9-bp deletion	HAPLOGROUP
Malay 6	126, 231, 311	-	+			+			Y
Malay 40	177, 223, 263, 266, 274, 311, 343	-	+			+			Y
Malay 55	168, 295, 296, 304	-	+			-	+		R21
Malay 14	295, 296, 304, 354	-	+			-	+		R21
Malay 86	166, 266, 304, 311	-	-						R1
Malay 98	145, 266, 304, 309, 325, 356	-	-						R
Malay 20	172, 193, 223, 344	-	-						R
Malay 93	176, 266, 357	-	-						R
Malay 29	192	-	-						R
Malay 59	209, 298, 355, 362	-	-	+					R
Malay 49	71, 172, 192A, 207, 325	-	-						R
Malay 41	93, 157, 201	-	-						R
Malay 4	95, 129, 140, 223, 265C, 271	-	-			+			R
Malay 45	136, 223, 257A, 261, 292, 294	-	-						N9a
Malay 21	223, 257A, 261, 292, 294	-	-						N9a
Malay 100	223, 257A, 261, 292, 294	-	-	+					N9a
Malay 92	183C, 189, 223, 311, 362	+	+	-					M-D
Malay 83	223, 271, 287, 319, 356, 362	+	+	-					M-D
Malay 61	189, 223, 298, 327, 399	+	+	+	+				M-C
Malay 65	223, 295, 362	+	+	+					M7c1c
Malay 71	223, 295, 362	+	+	+					M7c1c
Malay 102	223, 295, 362	+	+	+					M7c1c
Malay 85	129, 144, 148, 162, 192, 223, 241, 249, 265C, 311, 343	+	+						M4

SAMPLE	HAPLOTYPES	10397 *Alu* I	10394 *Dde* I	5176 *Alu* I	13262 *Alu* I	1715 *Dde* I	12406 *Hinc* II	9-bp deletion	HAPLOGROUP
Malay 26	129, 189, 192, 223, 297	+	+						M4
Malay 38	129, 189, 192, 223, 297	+	+						M4
Malay 35	129, 223, 290, 311, 325	+	+						M4
Malay 103	129, 172, 223, 256, 305, 309	+	+			+			M4
Malay 62	93, 129, 223, 234, 286, 290, 311, 362	+	+	+					M4
Malay 57	129, 144, 148, 223, 241, 255, 265C, 300, 311, 343	+	+						M3
Malay 31	179, 223, 264, 311	+	+						M3
Malay 91	223, 234, 300, 311	+	+						M3
Malay 66	223, 261, 311, 362, 390	+	+	+					M3
Malay 18	223, 311	+	+						M3
Malay 50	223, 311	+	+						M3
Malay 3	129, 140, 223, 265C, 271	+	+						M21a
Malay 99	93, 129, 223, 256, 271	+	+						M21a
Malay 42	93, 129, 223, 256, 271, 362	+	+	+					M21a
Malay 72	93, 129, 223, 256, 271, 362	+	+	+					M21a
Malay 75	93, 129, 223, 256, 271, 362	+	+	+					M21a
Malay 101	93, 129, 223, 256, 271, 362	+	+	+					M21a
Malay 78	209, 223, 224, 263, 278, 319	+	+						M2
Malay 10	51, 209, 223, 224, 263, 278, 319	+	+						M2
Malay 60	93, 209, 223, 224, 263, 278, 319	+	+						M2
Malay 88	93, 209, 223, 224, 263, 278, 319	+	+						M2
Malay 97	129, 192, 223, 297	+	+						M
Malay 106	145, 181, 192, 223, 291, 304	+	+	+			+		M
Malay 76	223	+	+		-				M
Malay 90	223, 234, 249, 261, 290, 360, 399	+	+	+					M
Malay 80	223, 234, 261, 290	+	+	+	-				M
Malay 9	223, 239, 289, 356	+	+						M

SAMPLE	HAPLOTYPES	10397 *Alu* I	10394 *Dde* I	5176 *Alu* I	13262 *Alu* I	1715 *Dde* I	12406 *Hinc* II	9-bp deletion	HAPLOGROUP
Malay 28	223, 261, 295	+	+	+	-				M
Malay 7	223, 295, 362	+	+	+					M
Malay 23	223, 295, 362	+	+	+					M
Malay 19	86, 129, 209, 223, 272	+	+						M
Malay 96	86, 129, 209, 223, 272	+	+						M
Malay 27	86, 192, 223, 297	+	+		-				M
Malay 12	93, 169, 184A, 223, 278	+	+	+					M
Malay 13	93, 223, 256, 274	+	+		-				M
Malay 30	Ref seq.	+	+	+					M
Malay 33	108, 129, 162, 172, 304	-	-				-		F1a1a
Malay 52	108, 129, 162, 172, 304	-	-				-		F1a1a
Malay 58	108, 129, 162, 172, 304	-	-				-		F1a1a
Malay 87	108, 129, 162, 172, 304	-	-				-		F1a1a
Malay 107	108, 129, 162, 172, 304	-	-				-		F1a1a
Malay 70	108, 129, 162, 172, 304, 311	-	-				-		F1a1a
Malay 11	108, 129, 162, 172, 304, 398	-	-				-		F1a1a
Malay 37	92, 108, 129, 162, 172, 234, 299, 304	-	-				-		F1a1a
Malay 56	129, 162, 172, 224, 304	-	-			+	-		F1a1
Malay 34	129, 162, 172, 304	-	-				-		F1a1
Malay 36	129, 162, 172, 304	-	-				-		F1a1
Malay 51	129, 162, 172, 304	-	-				-		F1a1
Malay 48	129, 172, 304	-	-				-		F1a
Malay 95	129, 172, 304	-	-				-		F1a
Malay 68	129, 172, 304	-	-				-		F1a
Malay 46	129, 172, 304, 311	-	-				-		F1a
Malay 25	129, 172, 304, 311	-	-				-		F1a
Malay 39	223, 390	+	+	+	-				E1

SAMPLE	HAPLOTYPES	10397 *Alu* I	10394 *Dde* I	5176 *Alu* I	13262 *Alu* I	1715 *Dde* I	12406 *Hinc* II	9-bp deletion	HAPLOGROUP
Malay 82	223, 362, 390	+	+	+					E1
Malay 24	51, 86, 148, 185, 223, 362, 390	+	+	+	-				E1
Malay 1	223, 291, 362, 390	+	+	+					E1
Malay 79	223, 261, 294, 362, 390	+	+	+					E1
Malay 53	223, 261, 362, 390	+	+	+	-				E1
Malay 84	223, 261, 362, 390	+	+	+	-				E1
Malay 105	140, 183C, 189, 243, 311	-	+					1	B5b
Malay 22	124, 140, 182C, 183C, 189, 261, 266A, 390	-	+			+		1	B5a
Malay 54	140, 183C, 189, 261, 266A, 269	-	+			+		1	B5a
Malay 47	140, 183C, 189, 266A	-	+					1	B5a
Malay 69	140, 183C, 189, 266A	-	+					1	B5a
Malay 74	140, 183C, 189, 266A	-	+			+		1	B5a
Malay 77	140, 183C, 189, 266A	-	+					1	B5a
Malay 89	140, 183C, 189, 266A	-	+					1	B5a
Malay 94	140, 183C, 189, 266A	-	+					1	B5a
Malay 44	140, 183C, 189, 266A, 304	-	+					1	B5a
Malay 5	140, 183C, 189, 266A, 362	-	+	+		+		1	B5a
Malay 104	136, 183C, 189, 217	-	-					1	B4b1
Malay 63	129, 154, 182C, 183C, 189, 217, 261	-	-					1	B4a
Malay 81	140, 182C, 183C, 189, 217, 274, 291	-	-					1	B4
Malay 17	140, 182C, 183C, 189, 217, 274, 335	-	-					1	B4
Malay 2	147, 183C, 184A, 189, 217, 235, 261	-	-					1	B4
Malay 43	147, 183C, 189, 217, 235	-	-					1	B4
Malay 73	147, 183C, 189, 217, 235	-	-					1	B4
Malay 15	182C, 183C, 189, 217, 274, 335	-	-					1	B4

A plus sign (+) denotes the site gain while the minus sign (-) indicates the site loss of the tested RFLP marker. '1' denotes the occurrence of 9-bp deletion

Table 5.4: RFLP polymorphisms, HVI sequence variance and mtDNA haplogroups of each Orang Asli samples.

SAMPLE	HAPLOTYPES	10397 *Alu* I	10394 *Dde* I	5176 *Alu* I	13262 *Alu* I	1715 *Dde* I	12406 *Hinc* II	9-bp deletion	HAPLOGROUP	TRIBE
OA 6	140, 183C, 189, 243T/C, 266A	-	+					1	B5a	Kinsiu
OA 10	140, 183C, 189, 266A	-	+					1	B5a	Kinsiu
OA 11	140, 183C, 189, 243T/C, 266A	-	+					1	B5a	Kinsiu
OA 24	140, 183C, 189, 243T/C, 266A	-	+					1	B5a	Kinsiu
OA 26	210, 223, 295, 311, 362	+	+	+					M	Kinsiu
OA 1	129, 223, 256, 271,362	+	+	+					M21a	Kinsiu
OA 4	93, 129, 223, 256, 271, 362	+	+	+					M21a	Kinsiu
OA 5	93, 129, 223, 256, 271, 362	+	+	+					M21a	Kinsiu
OA 7	93, 129, 223, 256, 271, 362	+	+	+					M21a	Kinsiu
OA 9	93, 129, 223, 224, 256, 271, 362	+	+	+					M21a	Kinsiu
OA 12	93, 129, 223, 256, 271, 362	+	+	+					M21a	Kinsiu
OA 13	129, 223, 256, 271,362	+	+	+					M21a	Kinsiu
OA 16	93, 129, 223, 256, 271, 362	+	+	+					M21a	Kinsiu
OA 17	93, 129, 223, 240, 256, 271, 362	+	+	+					M21a	Kinsiu
OA 18	93, 129, 223, 256, 271, 362	+	+	+					M21a	Kinsiu
OA 23	93, 129, 223, 256, 271, 362	+	+	+					M21a	Kinsiu
OA 27	93, 129, 223, 240, 256, 271, 362	+	+	+					M21a	Kinsiu
OA 28	93, 129, 223, 256, 271, 362	+	+	+					M21a	Kinsiu
OA 30	93, 129, 223, 256, 271, 362	+	+	+					M21a	Jahai
OA 34	93, 129, 223, 256, 271, 362	+	+	+					M21a	Jahai
OA 36	93, 129, 223, 256, 271, 362	+	+	+					M21a	Jahai
OA 37	93, 129, 223, 256, 271, 362	+	+	+					M21a	Jahai
OA 43	93, 129, 223, 256, 271, 362	+	+	+					M21a	Jahai

SAMPLE	HAPLOTYPES	10397 *Alu* I	10394 *Dde* I	5176 *Alu* I	13262 *Alu* I	1715 *Dde* I	12406 *Hinc* II	9-bp deletion	HAPLOGROUP	TRIBE
OA 50	93, 129, 223, 256, 271, 362	+	+	+					M21a	Jahai
OA 51	93, 129, 223, 256, 271, 362	+	+	+					M21a	Jahai
OA 31	93, 129, 223, 263, 381	+	+						M21b	Jahai
OA 15	93, 129, 209, 223, 261, 278, 325	+	+						M4	Kinsiu
OA 21	93, 129, 209, 223, 261, 278, 325	+	+						M4	Kinsiu
OA 32	223, 257A, 261, 292, 294	-	-						N9a	Jahai
OA 52	223, 257A, 261, 292, 294	-	-						N9a	Jahai
OA 2	168, 295, 296, 304	-	+			-	+		R21	Kinsiu
OA 3	168, 295, 296, 304	-	+			-	+		R21	Kinsiu
OA 8	168, 295, 296, 304	-	+			-	+		R21	Kinsiu
OA 14	168, 295, 296, 304	-	+			-	+		R21	Kinsiu
OA 19	168, 295, 296, 304	-	+			-	+		R21	Kinsiu
OA 20	168, 295, 296, 304	-	+			-	+		R21	Kinsiu
OA 22	168, 295, 296, 304	-	+			-	+		R21	Kinsiu
OA 29	168, 295, 296, 304	-	+			-	+		R21	Kinsiu
OA 33	86, 168, 295, 296, 304	-	+			-	+		R21	Jahai
OA 35	86, 168, 295, 296, 304	-	+			-	+		R21	Jahai
OA 38	86, 168, 295, 296, 304	-	+			-	+		R21	Jahai
OA 39	86, 168, 295, 296, 304	-	+			-	+		R21	Jahai
OA 40	86, 168, 295, 296, 304	-	+			-	+		R21	Jahai
OA 41	86, 168, 295, 296, 304	-	+			-	+		R21	Jahai
OA 42	86, 168, 295, 296, 304	-	+			-	+		R21	Jahai
OA 44	86, 168, 295, 296, 304	-	+			-	+		R21	Jahai
OA 45	86, 168, 295, 296, 304	-	+			-	+		R21	Jahai
OA 47	86, 168, 295, 296, 304	-	+			-	+		R21	Jahai
OA 48	168, 295, 296, 304	-	+			-	+		R21	Jahai
OA 49	86, 168, 295, 296, 304	-	+			-	+		R21	Jahai

SAMPLE	HAPLOTYPES	10397 *Alu* I	10394 *Dde* I	5176 *Alu* I	13262 *Alu* I	1715 *Dde* I	12406 *Hinc* II	9-bp deletion	HAPLOGROUP	TRIBE
OA 53	86, 168, 295, 296, 304	-	+			-	+		R21	Jahai
OA 54	86, 168, 295, 296, 304	-	+			-	+		R21	Jahai
OA 55	86, 168, 295, 296, 304	-	+			-	+		R21	Jahai
OA 56	86, 168, 295, 296, 304	-	+			-	+		R21	Jahai
OA 57	86, 168, 295, 296, 304	-	+			-	+		R21	Jahai
OA 58	86, 129G/A, 168, 295, 296, 304	-	+			-	+		R21	Jahai
OA 59	86, 168, 295, 296, 304	-	+			-	+		R21	Jahai
OA 60	86, 168, 295, 296, 304	-	+			-	+		R21	Jahai

A plus sign (+) denotes the site gain while the minus sign (-) indicates the site loss of the tested RFLP marker. '1' denotes the occurrence of 9-bp deletion.

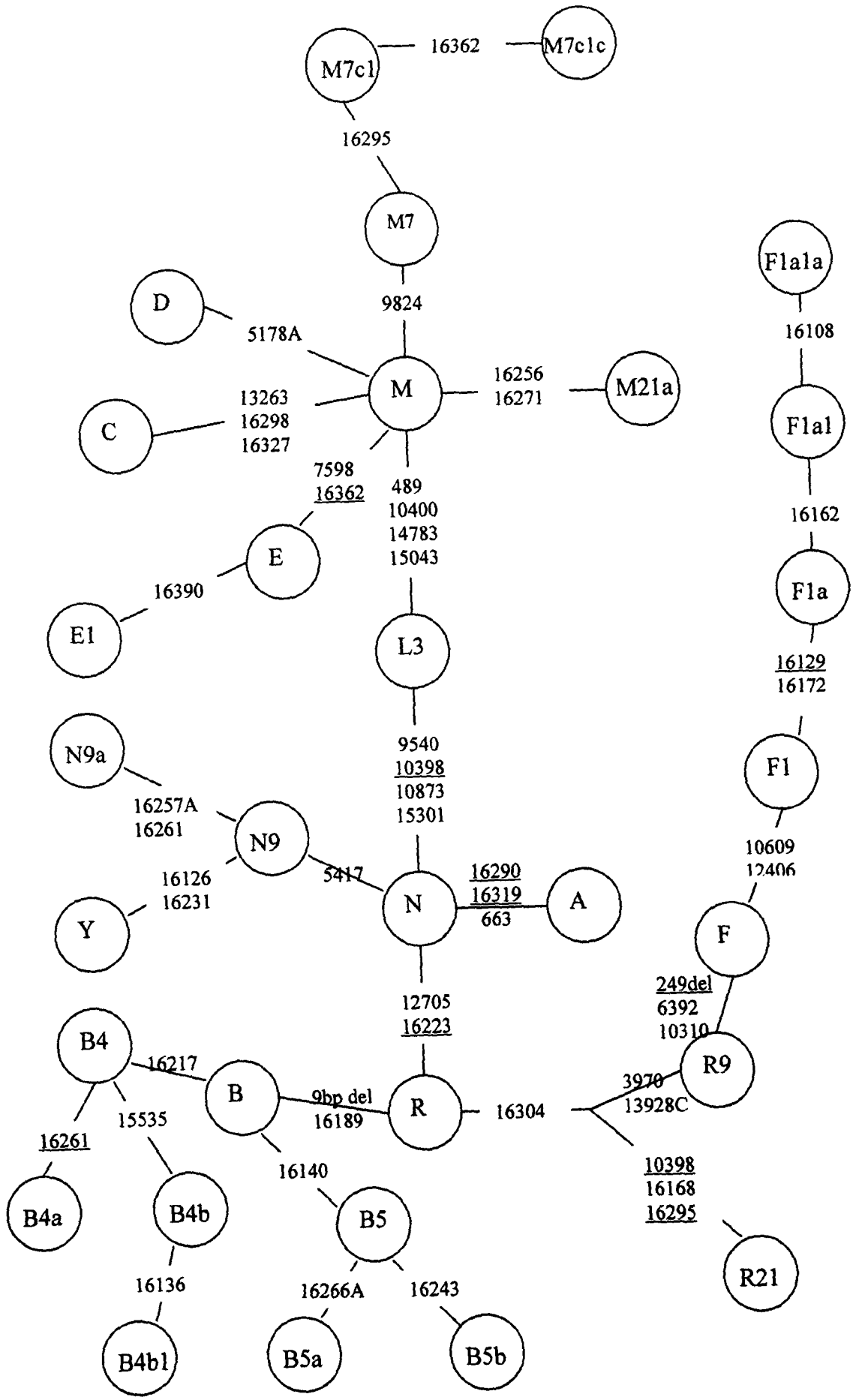
The overlapping *Alu* I and *Dde* I site gain at nps 10,394 and 10,397, which are prevalent in every Southeast Asian populations (Ballinger *et al.*, 1992; Schurr and Wallace, 2002), were found in high frequency within the Modern Malay and Orang Asli populations. This information had given the first impression of a common lineage shared between these two Peninsular Malaysia populations with other Southeast Asian populations. The Modern Malays, which are believed to be the descendants of proto-Malays mixed with modern Indians, Thai, Arab and Chinese, showed higher levels of diversity in their mtDNA haplogroup structure than the Orang Asli. This can be clearly seen by the distribution of mtDNA haplogroups and sub-clusters within the population. Sub-cluster C, D and E, known as typical 'Mongoloid' sub-clusters was observed at least once within the population. Haplogroup F, which was also suggested as a Mongoloid marker (Harihara *et al.*, 1988), was seen at a higher frequency in the population samples (16.67%). The 9-bp deletion, which nearly approached fixation in Polynesian (Melton *et al.*, 1995) and observed from low to moderate frequency in other Southeast Asian population (Schurr and Wallace, 2002), was found quite frequently in the Modern Malays (18.63%). Sub-cluster M7c1c, which is a characteristic of East Asian population (Kivisild *et al.*, 2002), was also observed within the population.

Having the 'Mongoloid' characteristic, Southeast Asian and East Asian specific haplogroups within the population, the Modern Malays at the same time showing some affinity to the Indian population. Several Indian-specific superhaplogroup M sub-clusters (M2, M3 and M4), which have not been highly represented in other Asian populations (Kivisild *et al.*, 1999), were observed between 4 to 6 times in the Modern Malays. The presence of several population specific mtDNA types in the Modern Malay population had demonstrated the intermixture of different lineages in this population.

The Orang Asli population had shown less diversity than the Modern Malays, which might be a result of genetic drift. Each of the Semang sub-group had only possessed 3 mtDNA types, with two (M21a and R21) of them were shared between these sub-groups. The Kinsiu showed high frequency of M21a while the Jahai exhibit higher frequency of R21. Sub-cluster M21a was also predominantly found in other Semang sub-groups studied elsewhere (Fucharoen *et al.*, 2001; Martin Richards, unpublished data). Haplogroup R21, which predominantly belongs to the Jahai, was only found twice within the Modern Malays. The loss of *Dde* I site at np position 1,715, which was predominant in the Jahai sub-group, was also found in the PNG highlands (Ballinger *et al.*, 1992) that resemble phenotypic similarity with the Semang. Studies done on the Malaysian and Philippines Semang had believed that they were basically the Australo-Melanesia inheritance, which extend westward from their core region around New Guinea (Bellwood, 1997) and being absorbed into the more numerous Mongoloid populations in Southeast Asia. The small number of mtDNA type shared between the Orang Asli and Modern Malay population suggesting limited relation of the ancestors for both populations. A schematic diagram showing all mtDNA haplogroups found in both populations, along with the defining mutations are shown in Figure 5.20.

The combination of RFLP analysis and HVI sequence information was proven to be the best tool for mtDNA classification. In future, with enough data available in the databases, information gained from RFLP analysis and sequence variation of mtDNA will not only be useful for human identification purposes, but can also be useful for determination of the origin of a perpetrator.

Figure 5.20: Schematic tree of mtDNA haplogroups found within the Modern Malay and Orang Asli populations of Peninsular Malaysia. The coding region polymorphisms and defining HVI motifs used for the haplogroups classification are shown along the connecting line. Recurrent mutations were underlined. Classifications of the haplogroups were based on information from Richards and Macaulay, 2000; Yao *et al.*, 2002 and Martin Richards, unpublished data.


M7c1
16362
M7c1c
16295
M7
9824
D
5178A
M
16256
16271
M21a
C
13263
16298
16327
7598
16362
489
10400
14783
15043
E
16390
E1
L3
9540
10398
10873
15301
N9a
16257A
16261
N9
16126
16231
Y
5417
N
16290
16319
663
A
12705
16223
B4
16217
B
9bp del
16189
R
16304
15535
16261
16140
B4a
B4b
B5
16136
16266A
16243
B4b1
B5a
B5b
F1a1a
16108
F1a1
16162
F1a
16129
16172
F1
10609
12406
F
249del
6392
10310
R9
3970
13928C
10398
16168
16295
R21

CHAPTER 6: MINISEQUENCING

6.1 Introduction

Single nucleotide polymorphisms had seen to have high potential as forensic DNA markers. SNPs within the coding region of mtDNA are very informative and being used to characterise mtDNA haplogroups in population studies by RFLP analysis (Hernstadt *et al.*, 2002). Although information from the hypervariable regions is more useful in forensic context, sequencing analysis of these regions is relatively time consuming and labour intensive especially when dealing with large amount of samples. Therefore, developing new mtDNA markers and rapid screening method for an effective exclusionary screening of forensic samples is very valuable.

SNPs within the mtDNA coding region were seen as the best candidate to be expanded as mtDNA markers. They were favoured as genetic marker due to their high abundance, low mutation rate and had greater potential for automation in the typing system (Erdogan *et al.*, 2001). A variety of SNP scoring system based on hybridisation (Wang *et al.*, 1998), ligation (Landegren *et al.*, 1988), polymerase (Syvanen *et al.*, 1990) and nuclease (Lee *et al.*, 1993; Lyamichev *et al.*, 1999) had been developed.

Minisequencing is one of the new growing techniques for interrogation of SNPs within the mitochondrial genome. This method was based on polymerase-mediated single base extension of oligonucleotide primer. Several minisequencing techniques such as the solid-phase fluorescent minisequencing (Tully *et al.*, 1996) and SNaPshot™ Multiplex Kit (Brandstatter *et al.*, 2003) have been used for analysis of mtDNA SNPs. The rapid

and discriminating nature of minisequencing make it well suited as a screening technique. Application of this technique in forensic cases could save a huge amount of work and time compared to the classical RFLP analysis since it allows multiplexing and high-throughput analysis. This method will also allows screening of non-definitive SNP markers, which are not amenable through restriction enzyme analysis.

6.1.1 SNaPshot™ Multiplex Kit

This kit was designed to interrogate up to 10 SNPs in a single tube. The chemistry of the reaction was based on a single nucleotide extension of an unlabelled primer. DNA polymerase contained in the reaction mix extended the primer by adding one fluorescently labelled ddNTPs at the 3′ end of the primer. This process was repeated in consecutive rounds of extension and termination to generate adequate amount of labelled product for analysis. To facilitate the detection of the labelled oligonucleotides, each ddNTPs was assigned with different dyes, which leads to a different peak colour on the analysed data (Table 6.1). The fluorescently labelled SNaPshot products are compatible for detection with the ABI PRISM® 310, 3100 and 3700 Genetic/DNA Analyzer System.

Table 6.1: Fluorescent dyes used to label each ddNTP in the SNaPshot™ Multiplex Ready Reaction Mix and the colour of each extension product.

ddNTP	Dye label	Colour of the analysed data
A	DR6G	Green
C	dTAMRA™	Black
G	dR110	Blue
T (U)	dROX™	Red

6.1.2 Selection of the Coding Region SNPs

The mtDNA coding region SNPs selected for this study were the one used as the defining substitution in haplogroups determination. These polymorphisms had been examined by restriction analysis by previous studies. All selected sites and the corresponding references are listed in Table 6.2.

6.2 Primer Design

The most important and vital step for minisequencing analysis was the primer design. Two sets of primers were designed prior to the analysis. The first set involved the amplification of the minisequencing templates and the second set were minisequencing primers used in the SNaPshot reaction. A total of 16 sets of primer were designed for amplification of minisequencing templates (Table 2.7) and 17 minisequencing primers were designed for the primer extension (Table 2.8). For the minisequencing primers, 14 of them were designed from the light strand while 3 primers were designed from the heavy strand (RA663G, RT8618C and RC10400T). Several standard considerations when designing primers such as GC content and non-complimentary 3' end were applied when designing both sets of primers. Extra considerations were taken when designing the minisequencing primers. Each minisequencing primer was designed to differ in length by at least 2 nucleotides from any of other primers to avoid the final products overlapping. When assigning the size for each primer, the colour of the expected SNaPshot product was taken into consideration. Peaks with the same colour that are too close to each other in size will be difficult to differentiate on the electropherogram.

All minisequencing primers were designed to have melting temperature of above 50 °C since the recommended annealing temperature for the SNaPshot control primer was 50 °C. The size of each minisequencing primer was adjusted to the required length by adding non-homologous polynucleotides tails at the 5′ end. Poly (dA) was chosen since it was among the non-homologous tails predicted to have minimal secondary structures (SNaPshot™ Multiplex Kit manual, Applied Biosystem). Type of non-homologous tails added to the minisequencing primer usually will not affect the signal pattern of the SNaPshot product, but poly (dT) might however interfere the addition of 3′ ddA.

Synthesis of long primers might result in heterogenous mixture of different molecular weight oligonucleotides. This will produce multiple signals from undesired primer extension and complicate the analysis of SNaPshot products. To avoid this problem, minisequencing primers with more than 30 nucleotides were requested to be HPLC purified. Another problem when dealing with primer extension method was the primer design did not offer any flexibility in regards to the 3′ end location. The only solution was to design the primer from the heavy strand when the sequence of the light strand was difficult to assay or produced a primer that might form hairpin structure or primer dimer. A low quality primer will reduce the efficiency of that particular primer extension and at the same time might also interrupt the efficiency of other primers in a multiplex reaction. In general, all minisequencing primers should be designed carefully in order to achieve a clear and reliable result when multiplexed in a single tube.

Table 6.2: mtDNA coding region SNPs selected for the minisequencing analysis and the established references.

Asian specific haplogroups		
Coding region SNPs	**Haplogroup**	**References**
663	A	Torroni *et al.*, 1992
13263	C	Torroni *et al.*, 1992
5178	D	Torroni *et al.*, 1992
7598	E	Torroni *et al.*, 1994
10398	M	Torroni *et al.*, 1993; Kivisild *et al.*, 2002
10400	M	Torroni *et al.*, 1993; Kivisild *et al.*, 2002
12406	F	Kivisild *et al.*, 1999
Caucasian specific haplogroups		
7028	H	Torroni and Wallace, 1994
14766	HV	Macaulay *et al.*, 1999; Derbeneva *et al.*, 2002
11719	preHV	Richards *et al.*, 2000
8618	*L3d	Rando *et al.*, 1998; Richards *et al.*, 2003
2352	*L3e	Rando *et al.*, 1998; Richards *et al.*, 2003
13368	T	Torroni *et al.*, 1996; Macaulay et al., 1999
12308	U	Torroni *et al.*, 1996; Macaulay *et al.*, 1999
4580	V	Torroni *et al.*, 1996; Macaulay *et al.*, 1999
14470	X	Macaulay *et al.*, 1999; Finnila *et al.*, 2001
8994	W	Torroni *et al.*, 1996; Macaulay *et al.*, 1999

*African specific haplogroups which also found in Caucasians (Herrnstadt *et al.*, 2002).

6.3 Results

6.3.1 Amplification and Purification of Minisequencing Templates

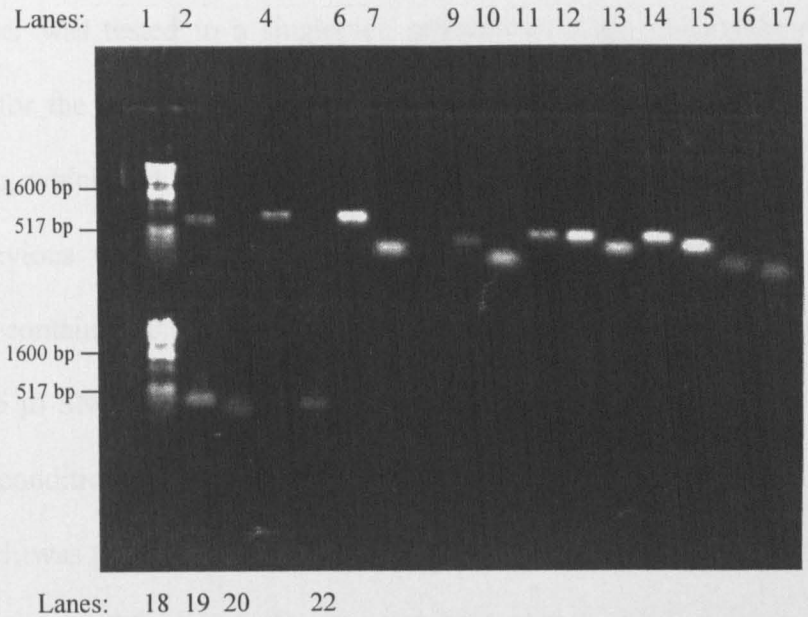
All 16 sets of primers and annealing temperature used to amplify the minisequencing templates are listed in Table 2.6. A total of 2 µl from the amplified products were electrophoresed in an EtBr stained agarose gel prior to the purification step. This was to make sure that the amplification process had successful and provide sufficient amount of minisequencing templates (Figure 6.1 and 6.2). All remaining products (15 µl) were purified by enzymatic reaction of SAP and *Exo* I. The SAP was used to dephosphorylate unincorporated dNTPs from the amplified products while the *Exo* I was used to degrade left over primers in the PCR reaction. Incomplete removal of dNTPs will allow them to participate in the primer extension reaction, leading to the formation of extraneous peaks and reduced the amplitude of the interest peaks. Left over PCR primers can also participate in the SNaPshot primer extension reaction and produce extraneous peaks, which then complicate the data interpretation. The purified templates were kept in 4 °C until further use.

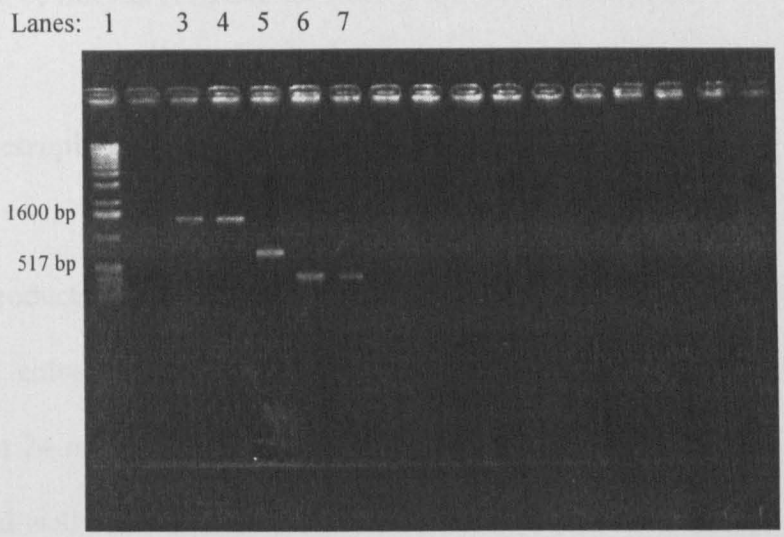
Figure 6.1: The amplification product for minisequencing templates.

Lane	Sample	Size (bp)
1	1 kb DNA ladder (1 μg/μl)	
2	Template for primer RA663G	710
4	Template for primer A13263G	834
6	Template for primer A13263G	834
7	Template for primer C5178A	322
9	Template for primer G12406A	440
10	Template for primer 7025C	226
11	Template for primer RT8618C	547
12	Template for primer T2352C	531
13	Template for primer G11719C	368
14	Template for primer C14766T	522
15	Template for primer G13368A	408
16	Template for primer A12308G	235
17	Template for primer G4580A	186
18	1 kb DNA ladder (1μg/μl)	
19	Template for primer G8994A	356
20	Template for primer T14470C	239
22	Template for primer A10398G and RC10400T	310

Figure 6.2: The amplification product for minisequencing templates.

Lane	Samples	Size (bp)
1	1 kb DNA ladder (1μg/μl)	
3	Template for primer G7598A	1517
4	Template for primer G7598A	1517
5	Template for primer RA663G	710
6	Template for primer G11719C	368
7	Template for primer G11719C	368





*2 µl of PCR products were electrophoresed on 1.2% agarose gel containing 1µl EtBr (10 mg/ml) for 45 min at 50 V using 1X TBE buffer.

6.3.2 SNaPshot Singleplex Reaction

Each primer was tested in a singleplex reaction in order to determine the optimum condition for the primer extension. When possible, samples representing each set of haplogroup, which had been determined by the HVI sequence data and RFLP analysis in the previous chapter, were used in the singleplex reaction. The initial primer extension contained 0.5 μM of the minisequencing primer, 0.5 μl minisequencing template, 5 μl SNaPshot Multiplex Ready Reaction Mix and 4 μl deionised water. The extension condition used for all primers was 96 °C for 10 s, 58 °C for 5 s and 60 °C for 30 s, which was performed for 25 cycles. After the extension process was completed, the remaining ddNTPs was dephosphorylated by adding 1 Unit of SAP and incubated at 37 °C for one hour. If left untreated, unincorporated ddNTPs will co-migrate with the fragment of interest and exhibits interference as extraneous peaks (appeared as peak more than 70 bp). Treatment with SAP had removed the 5′phosporyl groups and altered the migration of free ddNTPs and therefore prevented interference.

6.3.3 Electrophoresis of SNaPshot Product

SNaPshot products were resolved by electrophoresis on the ABI PRISM® 310 Genetic Analyser (5 colour system). The collection time on the GeneScan E5 module was shorten from 24 min to 18 min since the maximum expected peak was just 60 nt long and appeared well before the run ended. The injection time was also reduced from 5 s to 3 s to avoid overloading of the SNaPshot products. A matrix file for the Matrix Standard Set DS-02 was created before the first run of the SNaPshot products. This was done by running the Matrix Standard Set DS-02 (dR110, dRGG, dTAMRA, dROX,

LIZ) in the ABI PRISM® 310 Genetic Analyser and input the magnitude data obtained for each of the fluorescent dye peak into the matrix file creator (available on the ABI collection software/GeneScan® Analysis software). This matrix file was used along with the GeneScan E5 module in the ABI collection software for all SNaPshot products electrophoresis.

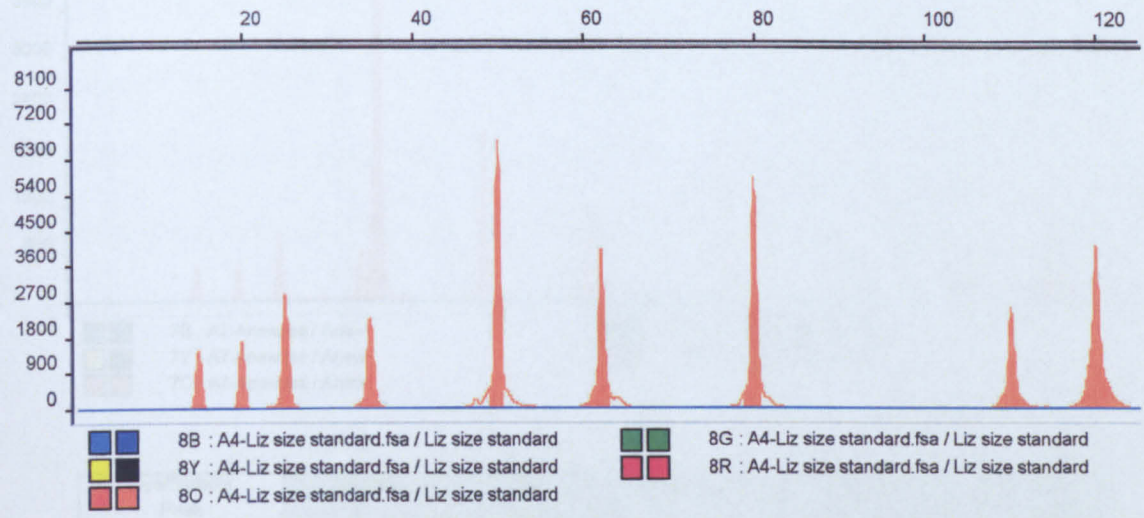
6.3.4 Analysis of Sample Files

The SNaPshot electrophoresis data was analysed using GeneScan® Analysis Software (version 3.1). GeneScan-120 LIZ size standard analysis parameter was used for automatic analysis and sizing of the SNaPshot products. The size for each peak of the GeneScan-120 LIZ size standard is shown in Figure 6.3.

6.3.5 Optimisation of the Singleplex Primer Extension

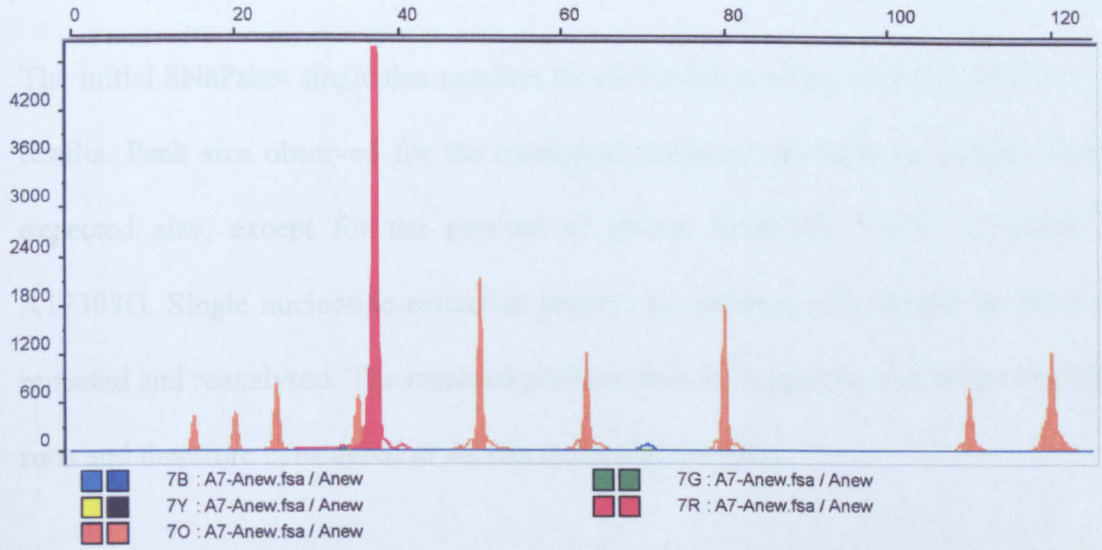
The standard reaction mix used for all singleplex reactions had resulted in very high signal for primer RA663G (Figure 6.4). A new reaction mix with a lower primer concentration was set up for this primer. The primer concentration was reduced from 0.5 μM to 0.2 μM. The same condition as the previous extension was used. Signal of the final extension product had significantly reduced when less primer concentration was used (Figure 6.5).

Figure 6.3: Electropherogram showing the size of each peak of the GeneScan-120 LIZ size standard used for automatic sizing of the SNaPshot products.



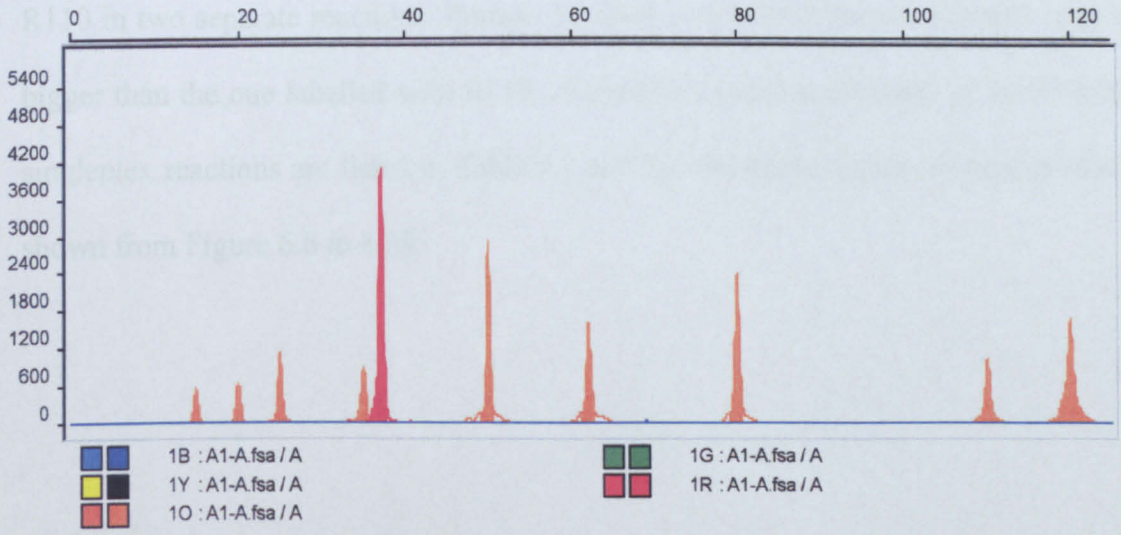
Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
8O, 1	10.34	15.00	1578	11609	2818
8O, 2	10.53	20.00	1813	14260	2870
8O, 3	10.72	25.00	3086	24401	2923
8O, 4	11.06	35.00	2478	19049	3014
8O, 7	11.58	50.00	7001	51151	3156
8O, 9	12.06	62.00	4303	37736	3287
8O, 12	12.66	80.00	6387	45378	3451
8O, 14	13.71	110.00	2739	26277	3739
8O, 15	14.05	120.00	4377	53350	3830

Figure 6.4: High signal when using 0.5 μM of primer concentration.



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
7R, 3	11.33	36.96	7599	60504	3088

Figure 6.5: Lower signal after primer concentration was reduced to 0.2 μM.



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
1R, 1	11.27	37.11	4537	30811	3073

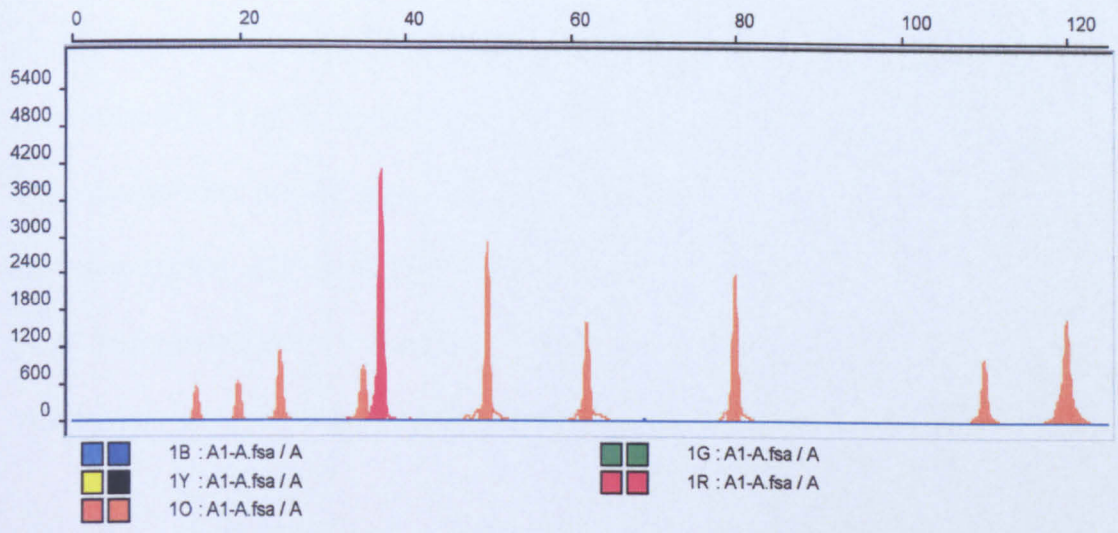
6.3.6 Analysis of Each Primer Extension Product

The initial SNaPshot singleplex reaction for all minisequencing primers had given good results. Peak size observed for the extension products was however smaller than the expected size, except for the product of primer RA663G, 7025C, RT8618C and A12308G. Single nucleotide extension process for primers with shorter products were repeated and reanalysed. The resolved product showed consistent size between different runs and therefore considered as correct extension product.

The difference between the expected and the observed size of the SNaPshot products was a result from incorporation of the fluorescent dyes, which have affected the mobility of the SNaPshot product and simultaneously altered the size of the product. The type of fluorescent dye incorporated to the primer had also affected product size. This scenario was observed when the same primer was incorporated by DR6G and R110 in two separate reactions. Product labelled with DR6G had consistently appeared bigger than the one labelled with R110. All primer extension products of the SNaPshot singleplex reactions are listed in Table 6.3 and the electropherogram of each product is shown from Figure 6.6 to 6.28.

Table 6.3: SNaPshot primer extension product obtained from each singleplex reaction of the minisequencing primer.

Primer	Haplogroup	Sample	Primer size (nt)	Product size (nt)	Peak colour	ddNTP added
RA663G	A	M4	36	37.11	Red	T
A13263G	C	M61	60	55.77	Blue	G
		M34		56.25	Green	A
C5178A	D	M83	56	53.70	Green	A
		M41		53.45	Black	C
G7598A	E	M1	40	39.30	Green	A
		M27		38.60	Blue	G
G12406A	F	M25	44	41.86	Green	A
		M83		40.38	Blue	G
RC10400T	M	M92	52	49.55	Green	A
		M15		48.78	Blue	G
A10398G	M	M92	48	45.97	Blue	G
		M15		46.72	Green	A
G11719C	preHV	M29	44	43.14	Green	A
14766C	HV	M41	32	31.98	Red	T
7028C	H	M30	40	40.85	Red	T
RT8618C	L3d	M66	56	56.49	Green	A
T2352C	L3e	M41	52	50.44	Red	T
G13368A	T	M20	48	46.26	Blue	G
A12308G	U	M29	36	37.04	Green	A
G4580A	V	M15	30	29.55	Blue	G
G8994A	W	M26	54	50.29	Blue	G
T14470C	X	M93	58	56.14	Red	T



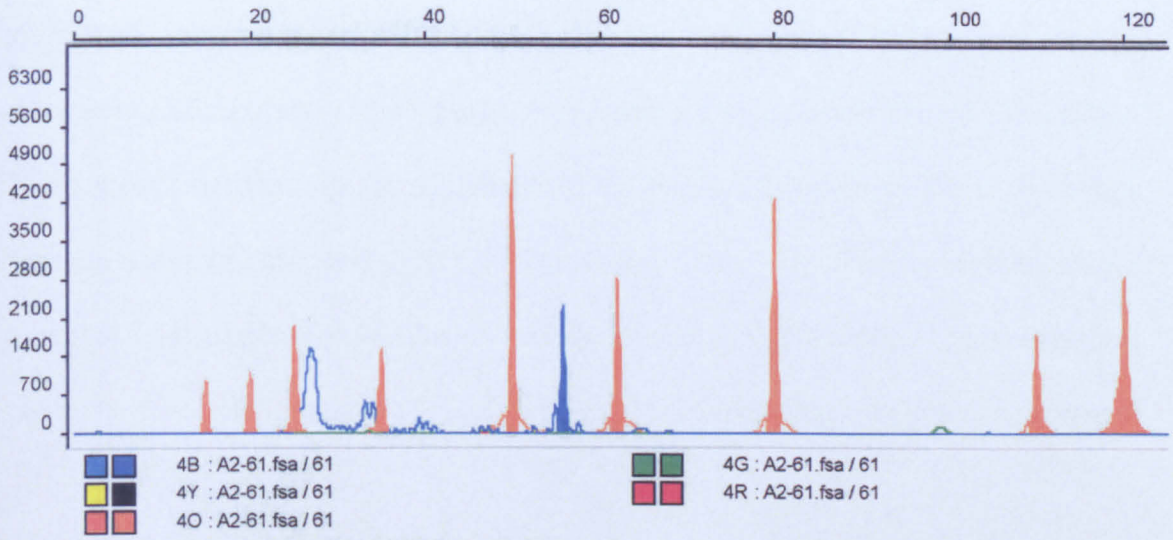
Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
1R, 1	11.27	37.11	4537	30811	3073

Figure 6.6: SNaPshot product of primer RA663G. This primer was designed for the screening of haplogroup A, which determined by an A to G transition at np 663. The primer was designed from the heavy strand since the light strand did not have a good GC balance (only 30% of GC content). The expected product would either be a red or a green peak, resulted from the addition of a ddTTP or ddCTP to primer RA663G. The sample used was M4 and the peak observed was red, indicating the addition of ddTTP. This result showed that M4 carried an A at np 663 and therefore did not belong to haplogroup A.

Figure 6.7: SNaPshot product for primer A13263G. This primer was design for screening of haplogroup C. It was defined by an *Alu* I site loss at np 13,262, created by an A to G transition at np 13,263. M61 was the only sample placed into haplogroup C by RFLP analysis and HVI sequence data. The result from this primer extension analysis had proved that M61 carried a nucleotide substitution at the interrogated site. A blue peak was observed, indicating an addition of ddGTP to the minisequencing primer.

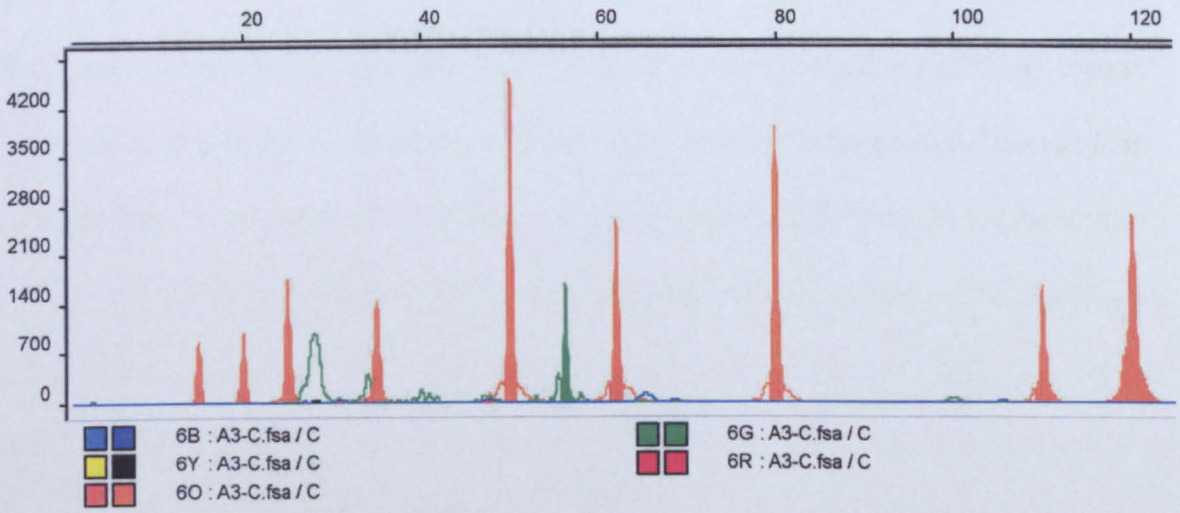
Figure 6.8: SNaPshot product for primer A13263G using a different sample from above. The sample used (M34) was not assigned to haplogroup C, either by the RFLP analysis or HVI sequence data. Result from this minisequencing had proved that this sample did not carry the nucleotide substitution (ddATP was added, green peak observed) and therefore did not belong to haplogroup C. Some extraneous peaks were observed before the actual peak appeared. In attempt to eliminate these peaks, minisequencing templates and SNaPhot products were purified again using SAP and *Exo* I. After several runs, the same profile was still observed. The possible explanation for the extraneous peaks would be heterogeneous mixture of the minisequencing primer. Since the extraneous peaks were much smaller and did not interfere with the actual peak, this result was accepted.

Figure 6.7:



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
4B, 22	12.68	55.77	2517	17395	3457

Figure 6.8:

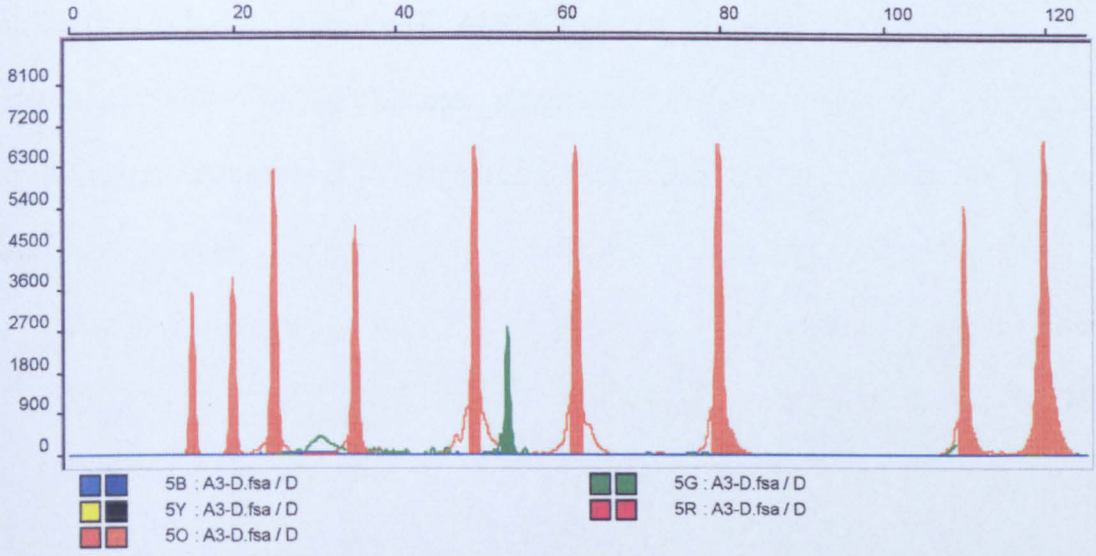


Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
6G, 19	12.72	56.25	1838	12950	3467

Figure 6.9: SNaPshot product for primer C5178A. The primer extension showed an addition of ddATP (green peak) for the sample used (M83). Nucleotide transversion from C to A had occurred at position 5,178, leading to an *Alu* I site loss at np 5,176. This characteristic had been used for classification of haplogroup D. The result obtained from this analysis confirmed that M83 belong to haplogroup D, as determined before by the HVI sequence and RFLP analysis.

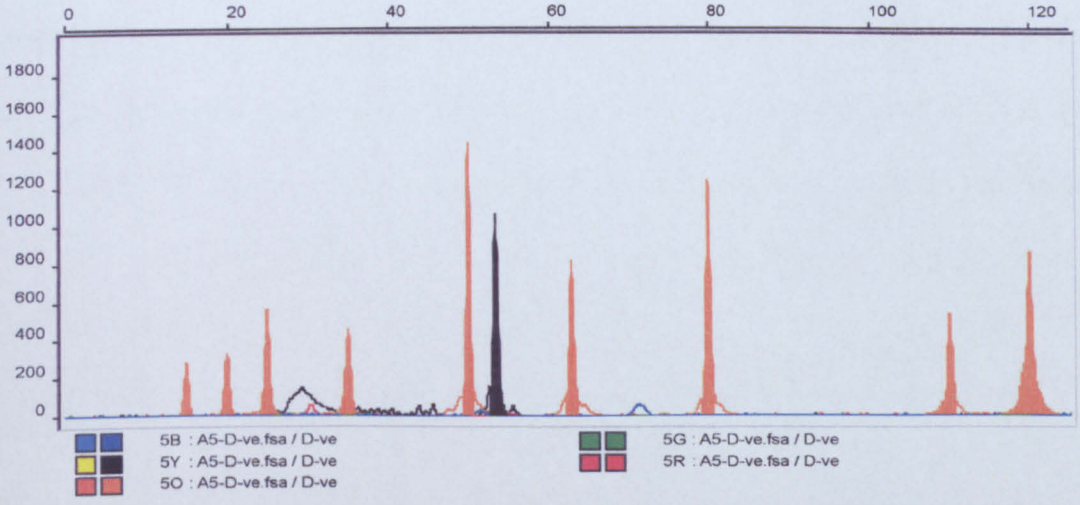
Figure 6.10: SNaPshot product for primer C5178A using a different sample from above (M41). A black peak indicates that a ddCTP was added to the minisequencing primer in the SNaPshot reaction. This result showed that nucleotide substitution at np 5,178 was not observed for sample M41 and therefore did not belong to haplogroup D.

Figure 6.9:



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
5G, 18	12.02	53.70	3054	20779	3276

Figure 6.10:

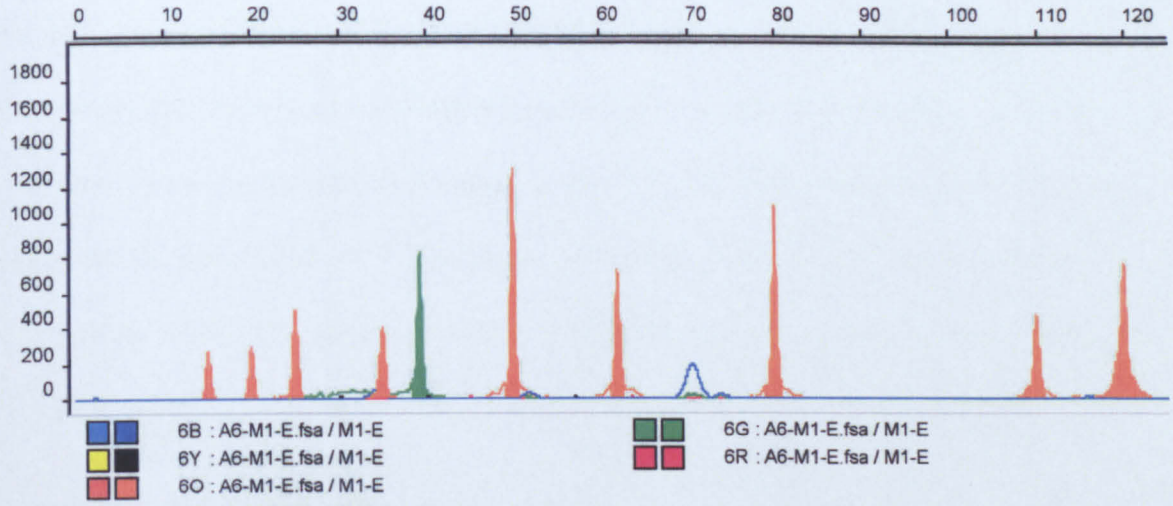


Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
5Y, 7	11.98	53.45	1163	8014	3266

Figure 6.11: SNaPshot product for primer G7598A. This primer was designed to assign samples into haplogroup E. It was characterised by *Hha*I site loss at np 7598, as a result from G to A transition at np position 7598. Sample used in the primer extension (M1) process carried the nucleotide substitution as a green peak was observed from the SNaPshot product analysis (ddATP was added). Thus, it belongs to haplogroup E.

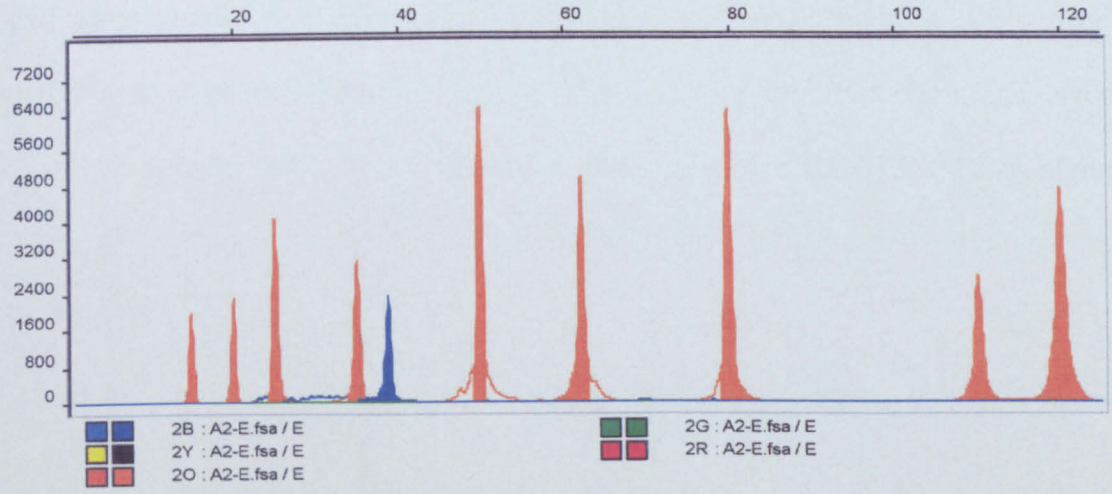
Figure 6.12: SNaPshot product for primer G7598A using a different sample from above. The product showed an addition of a ddGTP (blue peak), indicating that the sample tested (M27) did not belong to haplogroup E.

Figure 6.11:



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
6G, 4	11.21	39.30	929	6218	3055

Figure 6.12:

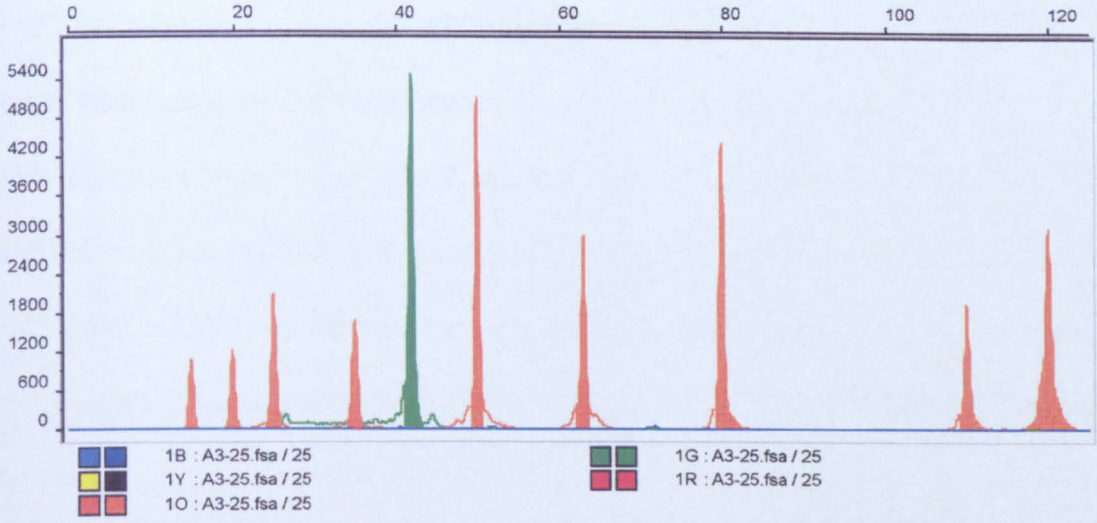


Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
2B, 11	10.91	38.60	2605	22735	2975

Figure 6.13: SNaPshot product for primer G12406A. This primer was designed for screening of haplogroup F. This haplogroup is characterised by G to A transition at np 12,406, leading to a *Hinc* II/*Hpa* I site loss at np 12,406. The sample tested with this primer (M25) was previously assigned into haplogroup F by RFLP analysis and HVI sequence data. Result from this SNaPshot analysis proved that M25 carried the nucleotide substitution (ddATP was added, green peak) and belongs to haplogroup F.

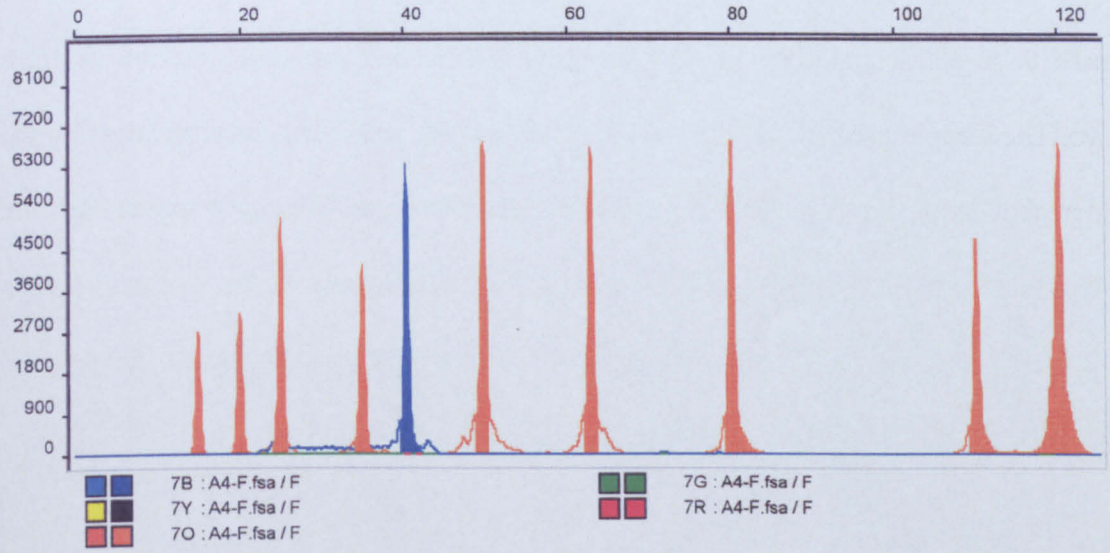
Figure 6.14: SNaPshot product for primer G12406A using a different sample from above. The sample used (M83) did not carry the nucleotide substitution as shown by the primer extension product. A blue peak observed in the electropherogram of the SNaPshot product indicates that a ddGTP was added to the primer.

Figure 13:



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
1G, 16	12.01	41.86	5952	43561	3274

Figure 14:

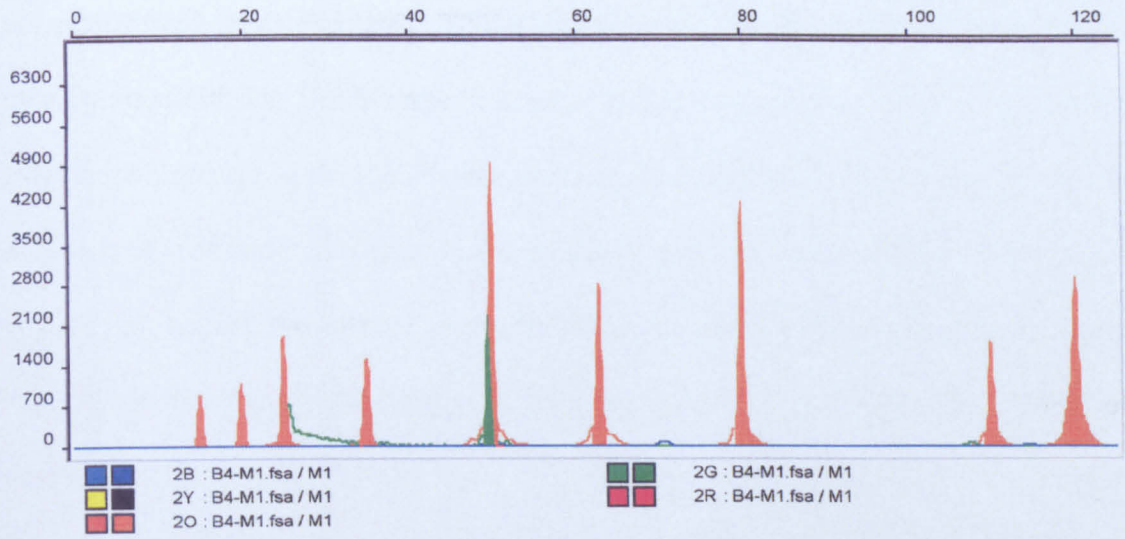


Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
7B, 18	11.60	40.38	7044	49533	3162

Figure 6.15: SNaPshot product for primer RC10400T. This primer was design from the heavy strand in order to avoid the neighbouring SNP at np 10398, which can complicate the multiplex analysis. This primer was used for screening of superhaplogroup M, determined by an *Alu* I site gain at np10397, created by a C to T transition at np 10400. Sample M92 used for this primer extension had been classified into haplogroup M by the RFLP analysis and the HVI sequence data. The SNaPshot product showed an addition of ddATP (complimentary to T in the light strand) to the minisequencing primer and this supported the result obtained from previous analyses.

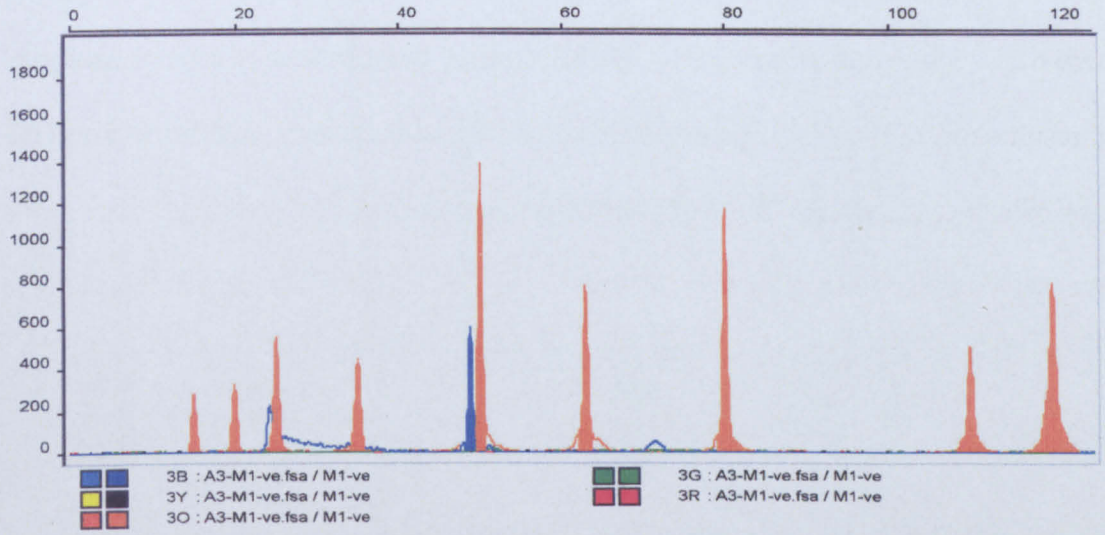
Figure 6.16: SNaPshot product using the same primer as above but different sample. This result showed that the sample used (M15) did not belong to superhaplogroup M. The primer was extended with a ddGTP (complimentary to C in the light strand), as showed by the blue peak. Note that there was a slight difference of the product size for the same primer. This was due to the effect of different dye incorporated to the primer, which then altered the mobility slightly (the sizes differ by just about 1 nucleotide).

Figure 15:



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
2G, 16	12.30	49.55	2540	16009	3353

Figure 16:

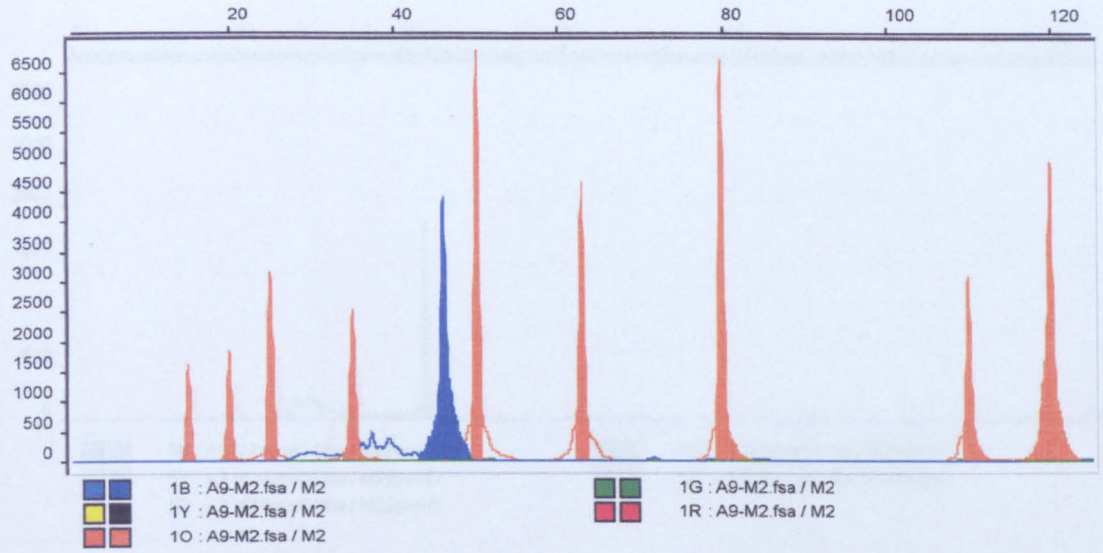


Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
3B, 7	11.92	48.78	658	4130	3249

Figure 6.17: SNaPshot product for primer A10398G. This primer was design to examine an A to G transition at np 10,398, which leads to a *Dde* I site gain at np 10,394. This mutation was always found along with the *Alu* I site gain at np position 10,397, which used for superhaplogroup M classification. Analysis of the SNaPshot product had shown an addition of ddGTP (blue peak) to primer A10398G and proved that sample M92 carried both *Alu* I and *Dde* I site gain at np 10,397 and 10,394, respectively.

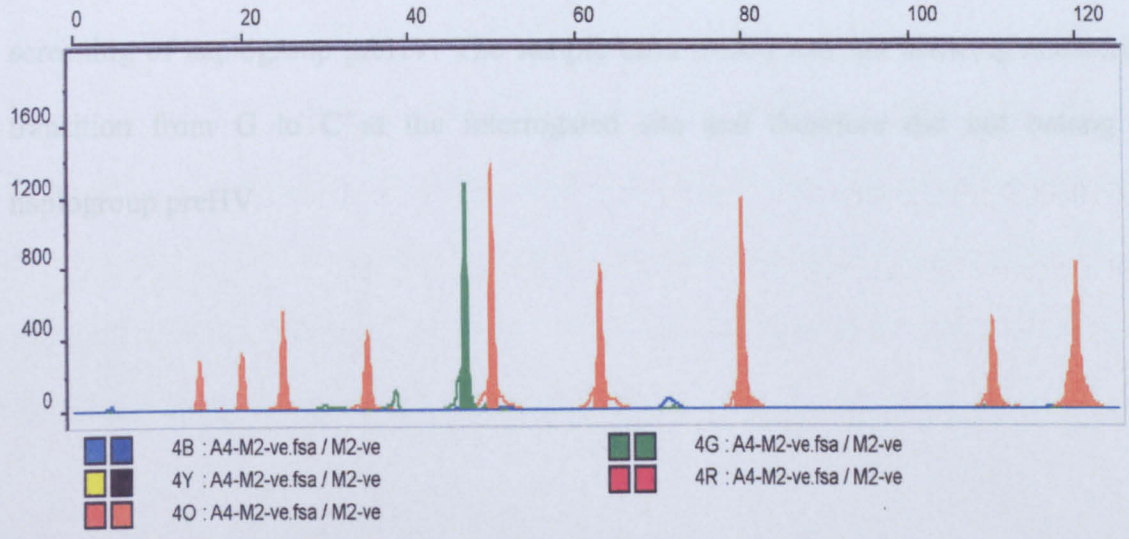
Figure 6.18: SNaPshot product for primer A10398G using sample M15. This sample was previously proved not belong to superhaplogroup M by primer extension using the RC10400T primer. A ddATP (green peak) was added to the primer, indicating that M15 did not carry the nucleotide substitution at the interrogated site.

Figure 17:

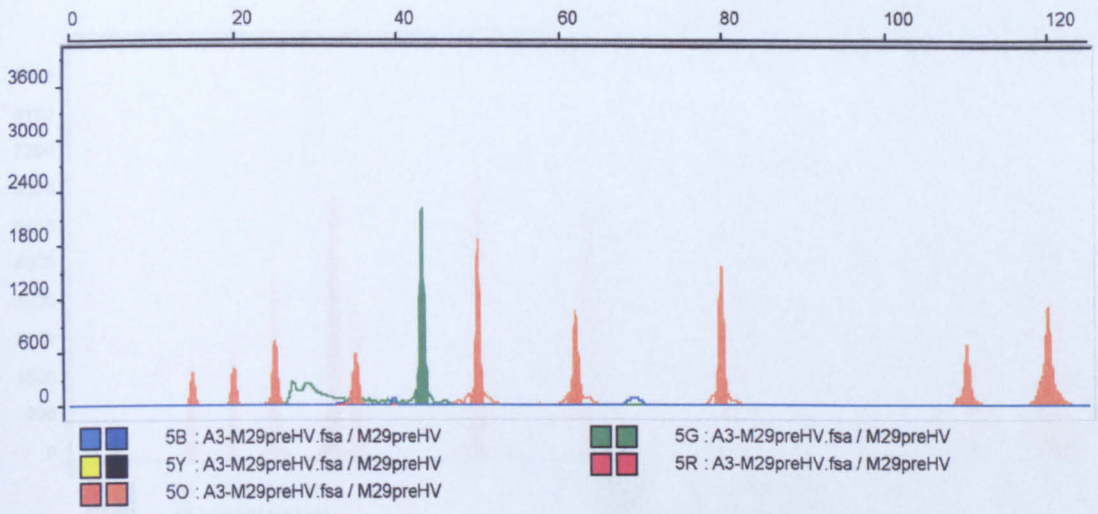


Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
1B, 10	12.06	45.97	4754	62314	3289

Figure 18:

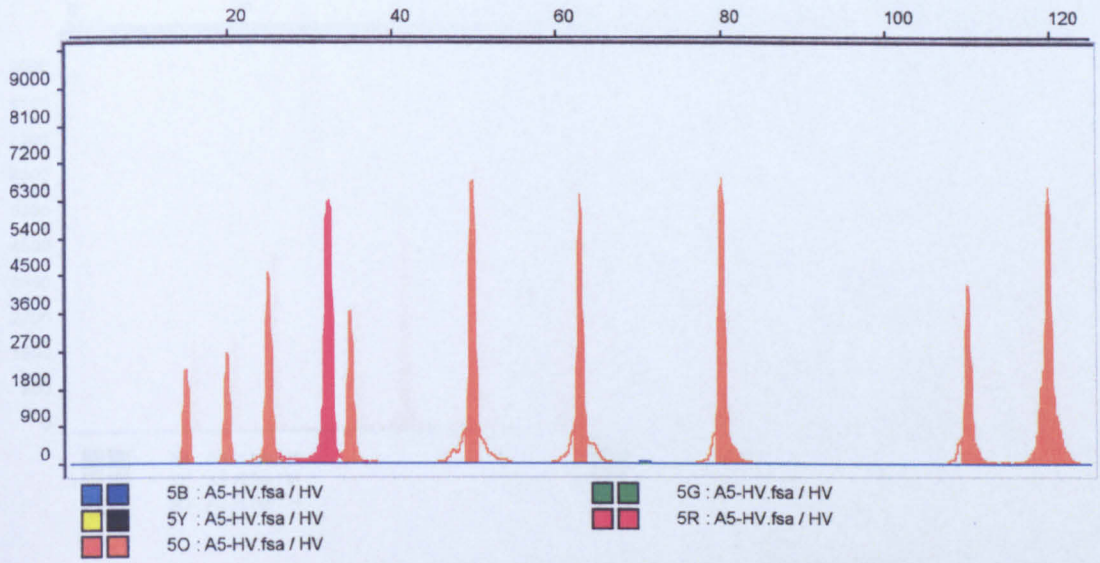


Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
4G, 3	11.80	46.72	1381	9376	3216



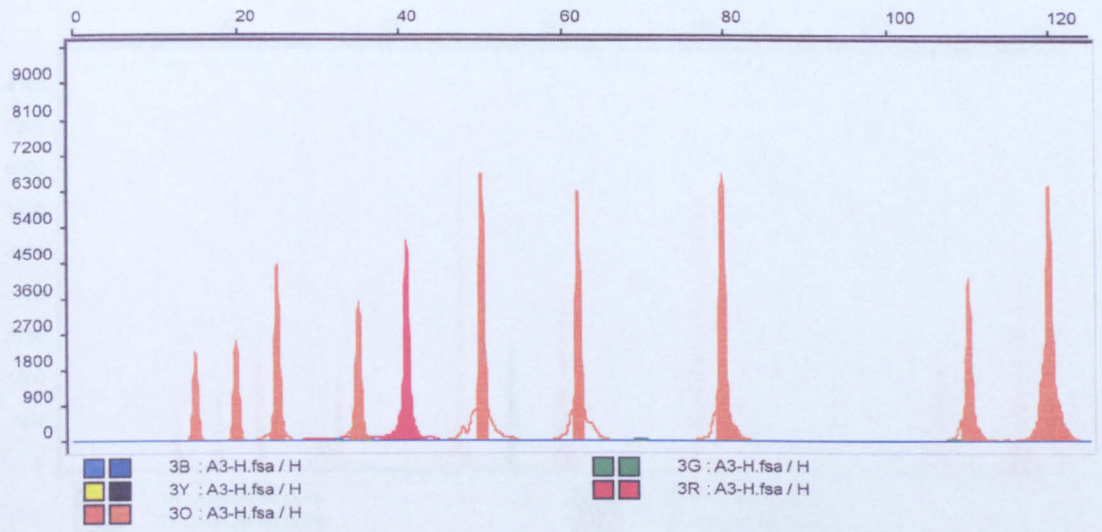
Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
5G, 13	11.44	43.14	2467	15770	3119

Figure 6.19: SNaPshot product for primer G11719C. This primer was designed for screening of haplogroup preHV. The sample used (M29) was not showing nucleotide transition from G to C at the interrogated site and therefore did not belong to haplogroup preHV.



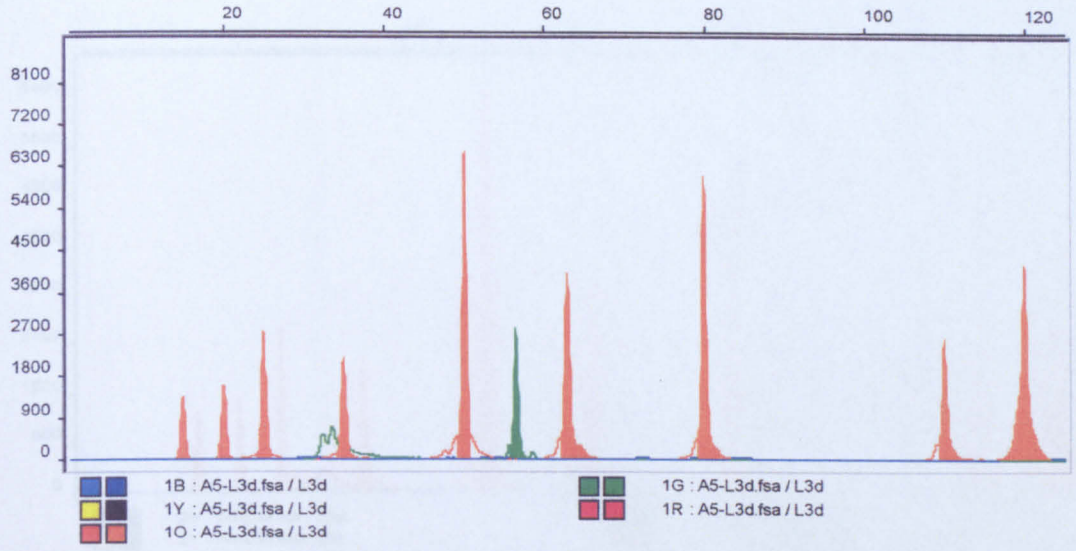
Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
5R, 2	11.25	31.98	6519	70425	3068

Figure 6.20: SNaPshot product for primer 14766C. Coding region sequence bearing a C at np 14,766 will lead to *Mse* I site loss at np 14,766, a characteristic for HV haplogroup classification. From the SNaPshot product analysis, polymorphism was detected at np 14,766 for sample M41. The minisequencing primer was extended with a ddTTP (red peak), indicating a transition from C to T had occurred. Sample M41 was therefore did not belong to haplogroup HV.



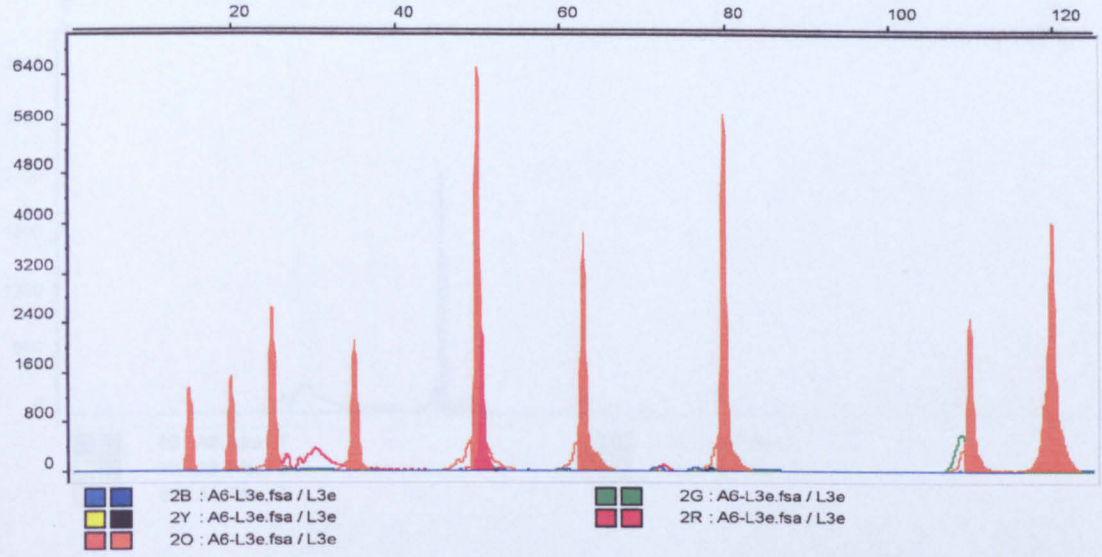
Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
3R, 9	11.70	40.85	5467	47760	3189

Figure 6.21: SNaPshot product for primer 7028C. The Cambridge Reference Sequence (Andrews *et al.*, 1999) carried a C at np 7,028 and this had simultaneously created an *Alu* I site loss at np 7,025, a characteristic to classify haplogroup H. The sample used (M30) in the primer extension reaction showed a transition from C to T at np 7,028, indicating that it did not belong to haplogroup H.



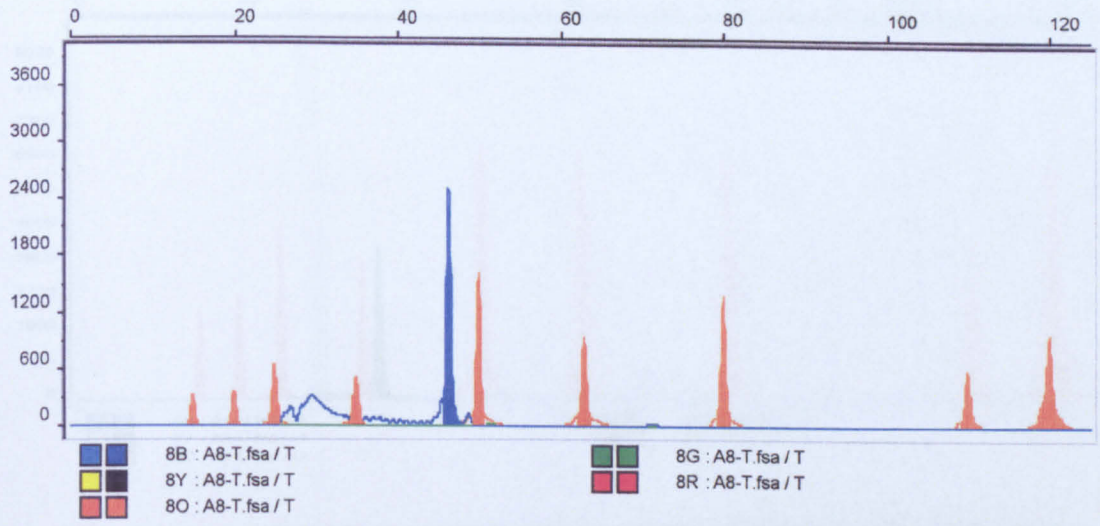
Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
1G, 10	12.93	56.49	3055	22654	3524

Figure 6.22: SNaPshot product for primer RT8618C. The primer was designed from the heavy strand in order to avoid a stretch of Cs ranging from np 8,605 to 8,609 on the light strand. It was used for screening of haplogroup L3d, characterised by *Mbo* I site loss at np 8,616. The SNaPshot product showed an addition of ddATP (green peak) to the minisequencing primer, indicating that the sample used (M66) carried an A at np 8,618 (for the reverse strand) and therefore did not belong to haplogroup L3d.



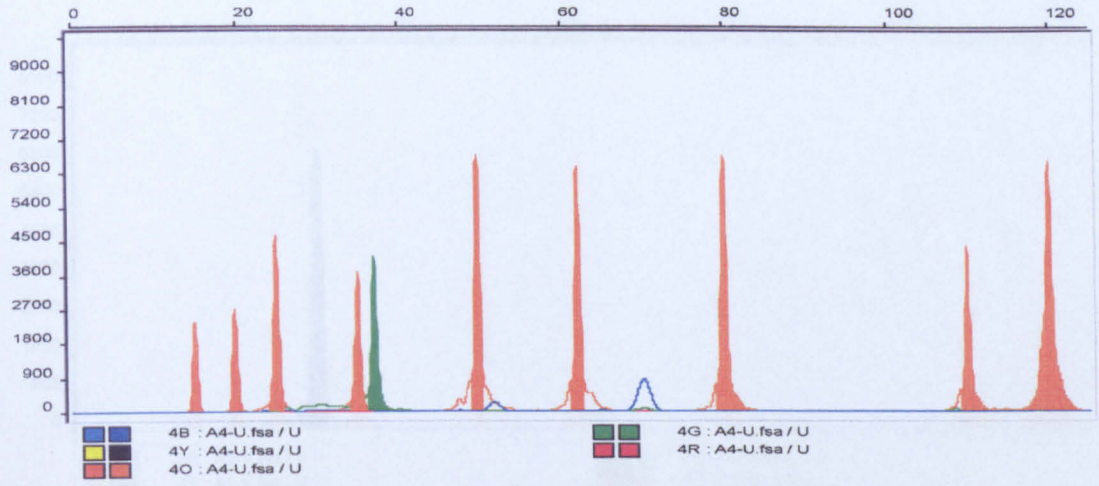
Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
2R, 15	12.49	50.44	2515	18901	3405

Figure 6.23: SNaPshot product of primer T2352C. This primer was designed for determination of T to C transition at np 2,352, which leads to an *Mbo* I site gain at np 2,349, a characterisation used for haplogroup L3e. Analysis of the SNaPshot product has shown an addition of ddTTP to the primer. The sample used (M41) did not carry the nucleotide substitution at np 2,352 and therefore did not belong to haplogroup L3e.



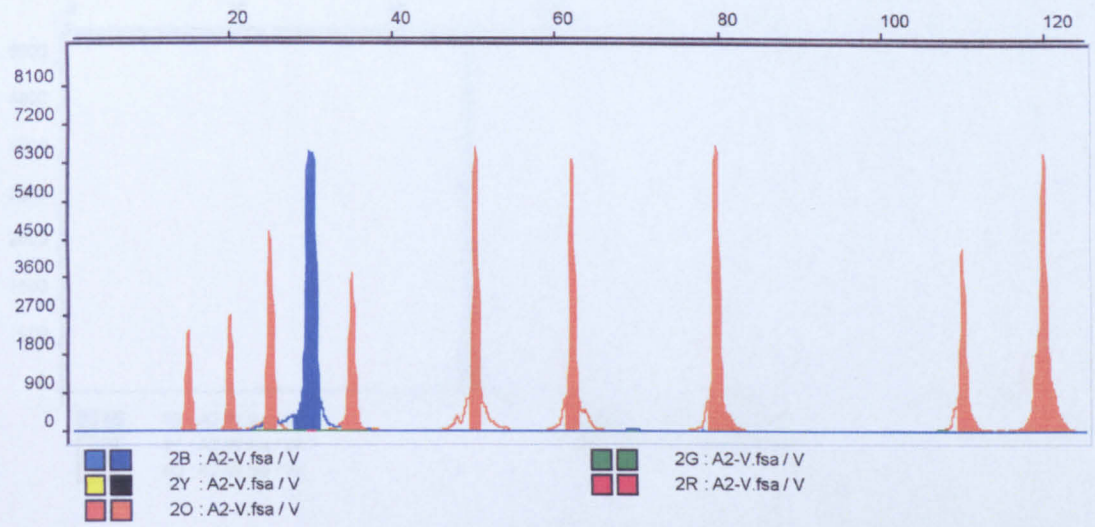
Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
8B, 16	11.63	46.26	2768	18360	3171

Figure 6.24: SNaPshot product for primer G13368A. The primer was design for screening of haplogroup T, defined by nucleotide transition from G to A at np 13,368, which then lead to *Ava* II/*Mbo* I site gain at np 13,367. Analysis of the SNaPshot product showed an addition of ddGTP to the minisequencing primer when tested with sample M93. This result indicates that this sample did not belong to haplogroup T.



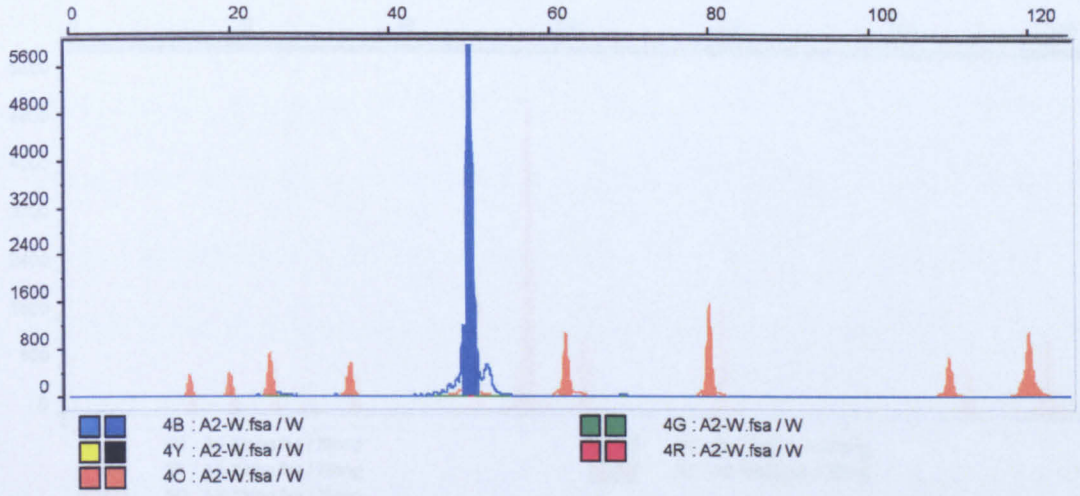
Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
4G, 7	11.52	37.04	4448	36924	3140

Figure 6.25: SNaPshot product for primer A12308G. This primer was designed to examine nucleotide transition from A to G at np 12,308, a characteristic for haplogroup U classification. An addition of ddATP (green peak) to the minisequencing primer was observed, indicating that the sample tested (M29) did not carry nucleotide substitution and therefore did not belong to haplogroup U.



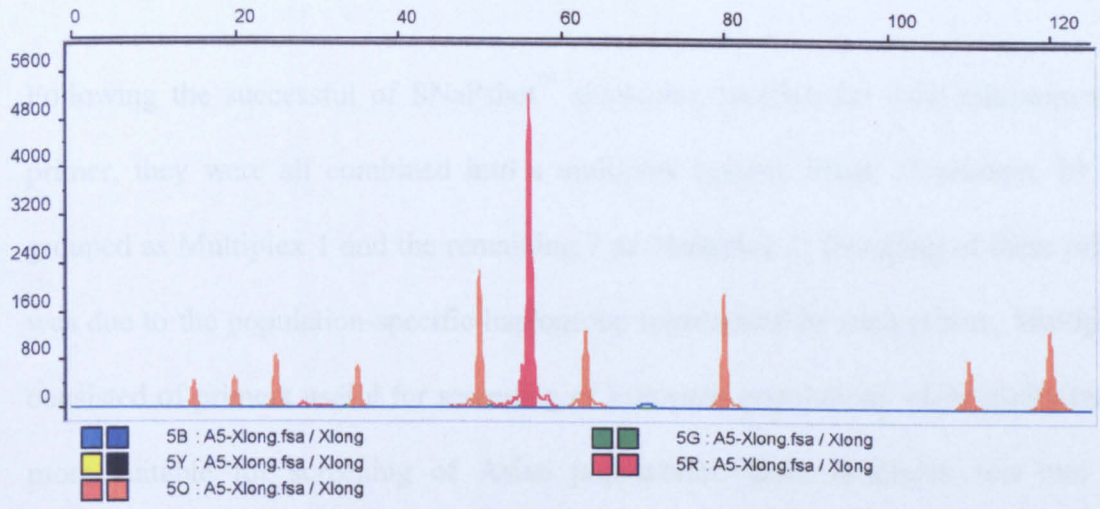
Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
2B, 5	11.37	29.55	6892	98172	3100

Figure 6.26: SNaPshot product for primer G4580A. This particular primer was designed to examine the existence of G to A transition at np 4,580, leading to *Nla* III site loss at np 4,577. The SNaPshot product produced a blue peak, showing an addition of a ddGTP to the minisequencing primer. The sample tested with this primer (M15) did not carry the polymorphism and therefore did not belong to haplogroup V.



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
4B, 6	11.74	50.27	7540	77992	3201

Figure 6.27: SNaPshot product for primer G8994A. This primer was design to interrogate transition from G to A at np 8,994, which leads to *Hae* III site loss at np 8,994. It was used to classify samples into haplogroup W. Analysis of the SNaPshot product showed that the sample used (M26) in this primer extension did not carry the substitution (blue peak, a ddGTP was added) and thus did not belong to haplogroup W.



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
5R, 7	11.97	56.14	5688	39526	3264

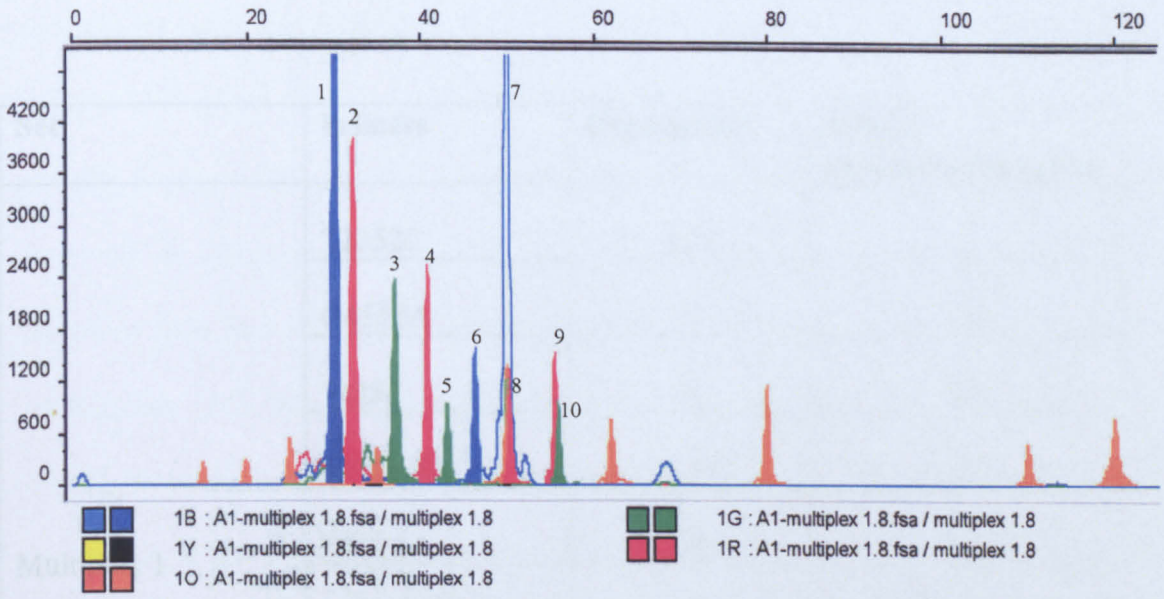
Figure 6.28: SNaPshot product for primer T14470C, designed for screening of haplogroup X. This haplogroup was defined by an *Acc* I site gain at np 14,465, created by T to C transition at np 14,470. The sample tested (M93) using this primer did not carry the polymorphism since the SNaPshot product showed a ddTTP extension to the minisequencing primer. M93 therefore did not belong to halogroup X.

6.3.7 SNaPshot Multiplex Reaction

Following the successful of SNaPshot™ singleplex reaction for each minisequencing primer, they were all combined into a multiplex system. From 17 primers, 10 were grouped as Multiplex 1 and the remaining 7 as Multiplex 2. Grouping of these primers was due to the population-specific haplogroup represented by each primer. Multiplex 1 consisted of primers useful for screening of European populations while Multiplex 2 is more suitable for screening of Asian populations. Both multiplex sets and their corresponding primers are listed in Table 2.8.

The starting concentration for each primer in the SNaPshot™ multiplex reaction was 1 μM. According to the SNaPshot™ Multiplex Kit manual (Applied Biosystem), successful results were obtained using primer concentration ranging between 0.05 μM to 1 μM in a six-primer mixture. Any high or low signal observed for each primer could be adjusted by increasing or decreasing the concentration of that particular primer. From the initial multiplex reaction, a very high signal was obtained for primer G4580A, 14766C and G8994A in Multiplex 1 (Figure 6.29) and a very low signal was observed for primer RC10400T in Multiplex 2 (Figure 6.30). All other primers in Multiplex 1 and Multiplex 2 showed good quality result. In order to lower the signal for primer G4580A, 14766C and G8994A, concentration for these primers was reduced to 0.1 μM for G4580A, 0.4 μM for 14766C and 0.2 μM for G8994A. Primer concentration for 7028C was also reduced to 0.6 μM since it also showed considerably high signal then other primers. Primer concentration for RC10400T was increased up to 2 μM in Multiplex 2 since the peak observed was very low. The optimised primer concentration for Multiplex 1 and Multiplex 2 is shown in Table 6.4.

Figure 6.29: Multiplex 1 showing high signal for peak 1, 2 and 7.



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
1B, 4	11.15	30.16	6777	48785	3040
1B, 10	11.75	46.40	1713	12339	3203
1B, 13	11.88	49.82	7453	66266	3240
1G, 5	11.39	37.08	2587	21498	3106
1G, 8	11.62	43.09	1052	7126	3168
1G, 12	12.13	55.88	1065	7250	3308
1R, 5	11.22	32.28	4384	31918	3060
1R, 11	11.54	40.89	2775	19727	3145
1R, 15	11.90	50.27	1511	9889	3245
1R, 18	12.12	55.61	1681	11055	3305

Peak number	Primer
1	G4580A
2	14766C
3	A12308G
4	7028C
5	G11719C
6	G13368A
7	G8994A
8	T2352C
9	T14470C
10	RT8618C

Table 6.4: Optimised concentration for each primer used in the multiplex SNaPshot reaction.

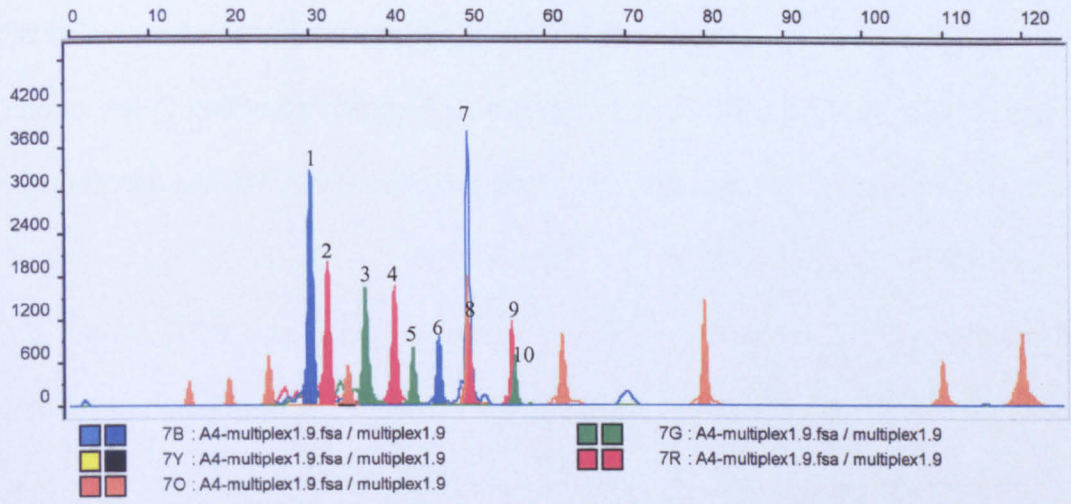
Set	Primers	Haplogroup	Primer concentration (μM)
Multiplex 1	T2352C	L3e	1
	G4580A	V	0.1
	7028C	H	0.6
	RT8618C	L3d	1
	G8994A	W	0.2
	G11719C	preHV	1
	A12308G	U	1
	G13368A	T	1
	T14470C	X	1
	14766C	HV	0.4
Multiplex 2	RA663G	A	1
	C5178A	D	1
	G7598A	E	1
	A10398G	M	1
	RC10400T	M	2
	G12406A	F	1
	A13263G	C	1

The optimised primer concentration for Multiplex 1 has given satisfactory results. The signal obtained for each peak was relatively well balanced. Figure 6.30 displays the electropherogram for Multiplex 1 primer extension products. From the analysed data, sample M41 tested with this multiplex system did not belong to any of the European-specific haplogroups. All 10 interrogated sites did not carry any of the polymorphism involved in characterisation of these haplogroups. The RFLP analysis and HVI sequence data had also failed to classify this sample into any of the Asian-specific haplogroup. M41 was therefore assigned into haplogroup R until further study could identify the exact haplogroup where it belongs. Recent report by Richards *et al.*, 2003, had indicated a RFLP site for haplogroup R classification. It was characterised by an *Mbo* II site gain at position 12,703. This site was however not tested for M41 due to time constrains. The main target was achieved by successful extension for all minisequencing primers involved in this multiplex set.

A good quality result was also observed for the optimised concentration of Multiplex 2. The signal for each peak was good, except for primer RC10400T where the signal obtained was low. The signal did not improve even when the primer concentration was increased. Further optimisation such as reducing and increasing the annealing and extension time/temperature, reducing other primers concentration and omitting the A10398G primer from the multiplex set (since they share the same template) were done. These optimisation techniques still did not improve the peak signal for primer RC10400T. This result might show the limitation of minisequencing technique, where some primers will not work in a multiplex system. From this primer extension result, the sample tested with Multiplex 2 (M83) obviously belongs to haplogroup D. This result matched the information gained from the RFLP analysis and HVI sequence data.

Figure 6.30: Electropherogram showing the optimised primer extension result of Multiplex 1. Ten distinct peaks were observed, indicating successful extension for all primers pooled in Multiplex 1. Size of each peak observed in this multiplex primer extension was similar to the size obtained from singleplex reaction. The overlapping peaks observed in this electropherogram (peak 7 and 8; peak 9 and 10) resulted from altered mobility of each primer due to the fluorescent dye incorporation. Since the colour of these overlapping peaks was different, the result was clearly seen and did not affect data interpretation.

Figure 6.30:

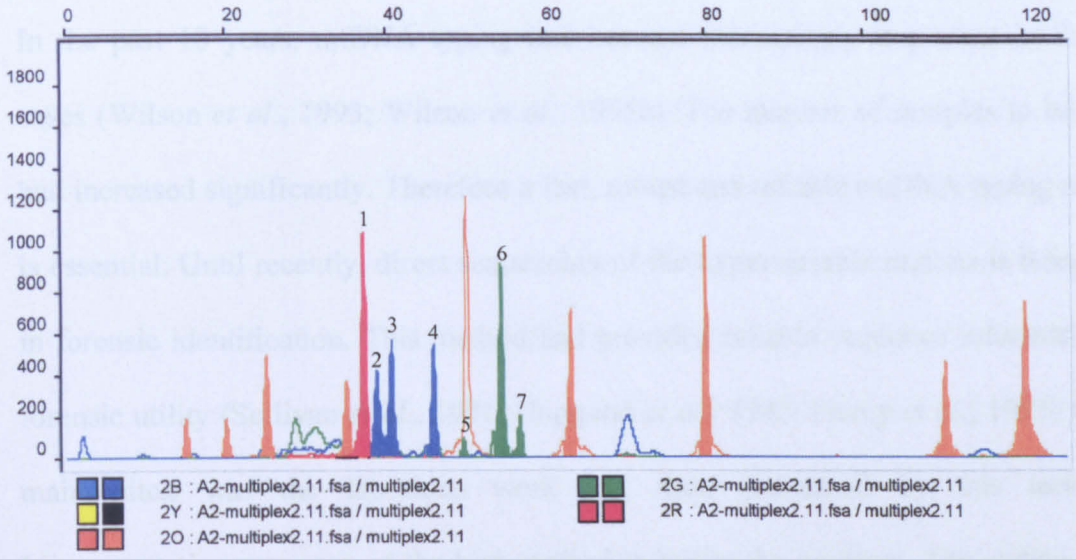


Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
7B, 4	**11.55**	**30.20**	**3494**	**27496**	**3149**
7B, 9	**12.17**	**46.41**	**1054**	**7639**	**3319**
7B, 12	12.32	50.08	4153	30153	3360
7G, 5	**11.81**	**37.11**	**1756**	**15962**	**3219**
7G, 8	**12.04**	**43.12**	**877**	**6286**	**3283**
7G, 10	**12.58**	**55.93**	**782**	**5618**	**3430**
7R, 4	**11.63**	**32.37**	**2144**	**17342**	**3171**
7R, 8	**11.95**	**40.79**	**1819**	**15122**	**3258**
7R, 11	**12.34**	**50.42**	**1223**	**9013**	**3364**
7R, 13	**12.57**	**55.68**	**1279**	**8934**	**3427**

Peak number	**Primer**	**Primer size**	**Peak colour**	**ddNTP added**	**Peak Size**
1	G4580A	30	Blue	G	30.20
2	14766C	32	Red	T	32.37
3	A12308G	36	Green	A	37.11
4	7028C	40	Red	T	40.79
5	G11719C	44	Green	A	43.12
6	G13368A	48	Blue	G	46.41
7	G8994A	54	Blue	G	50.08
8	T2352C	52	Red	T	50.42
9	T14470C	58	Red	T	55.68
10	RT8618C	56	Green	A	55.93

Figure 6.31: Electropherogram showing primer extension products for Multiplex 2. Primers pooled in this multiplex set produced significant peaks, except for primer RC10400T. Signal for this primer (peak 5) was too low, even after further optimisation. This result might show the limitation of minisequencing technique, where some primers will not work in a multiplex system.

Figure 6.31:



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
2B, 8	11.59	38.70	445	3829	3159
2B, 9	11.66	40.63	618	4650	3179
2B, 11	11.86	45.87	598	3895	3234
2G, 7	12.00	49.53	118	736	3273
2G, 9	12.19	54.29	1027	7039	3323
2G, 10	12.28	56.67	227	1670	3348
2R, 1	11.53	37.15	1179	8685	3143

Peak Number	Primer	Primer size	Peak colour	ddNTP added	Peak size
1	RA663G	36	Red	T	37.15
2	G7598A	40	Blue	G	38.70
3	G12406A	44	Blue	G	40.63
4	A10398G	48	Blue	G	45.87
5	RC10400T	52	Green	A	49.53
6	C5178A	56	Green	A	54.29
7	A13263G	60	Green	A	56.67

6.4 Discussion

In the past 10 years, mtDNA typing had become increasingly important in forensic cases (Wilson *et al.*, 1993; Wilson *et al.*, 1995a). The number of samples to be typed had increased significantly. Therefore a fast, robust and reliable mtDNA typing method is essential. Until recently, direct sequencing of the hypervariable regions is being used in forensic identification. This method had provided reliable sequence information for forensic utility (Sullivan *et al.*, 1991; Hopgood *et al.*, 1992; Piercy *et al.*, 1993) but the main hitch was the laborious work and time consumed by this technique. Minisequencing was seen as the best method to buffer the problem. The coding region SNPs were markers chosen to be developed along with this method.

The fact that the coding region is 15 times larger than the control region, a huge portion of mtDNA variation should occurred within this region. Previous study within the Caucasian population had shown the potential of the coding region SNPs in increasing the power of discrimination when a common HVI and HVII sequence was encountered (Parsons and Coble, 2001, Coble *et al.*, 2004).

This study had demonstrated that minisequencing was a straightforward and a robust method for screening of mtDNA coding region SNPs. The availability of ABI PRISM® SNaPshot™ Multiplex Kit in the market had facilitated this technique by providing all components required for the primer extension process in a single reaction mix. Sensitivity test had shown that minisequencing was powerful enough to produce informative result from as little as 1 ng of PCR product (Fiorentino *et al.*, 2003), which

often failed by direct sequencing. This fact will become very useful when dealing with limited or severely degraded forensic samples.

Another interesting feature of minisequencing was the ability for multiplexing. This had reduced the cost, time and effort for minisequencing. Possibility of cross contamination was also reduced since reaction was performed in a single tube. This study had successfully multiplexed up to ten primers in a single tube. The products were resolved in a single electrophoresis and the data analysis was fully automated. The computer-assisted analysis allowed exact base identity, which facilitate data interpretation and at the same time reduced the source of error. Due to the rapid nature of the minisequencing technique, it could be used as an effective exclusionary screening for forensic analysis. For samples that were not excluded from this preliminary screening, further analysis by direct sequencing can be performed.

In future, development of more multiplexes targeting mtDNA polymorphisms within the control and coding regions might reduce the matching probability of mtDNA type and increased the utility of mtDNA in forensic cases. If this had achieved, direct sequencing of the hypervariable regions and classical RFLP analysis for identification of coding region mutations would become unnecessary and consequently saved huge amount of work and time when dealing with mtDNA analysis.

CHAPTER 7: CONCLUSION

The past few years have seen the establishment of mtDNA profiling, with increased use and emphasis. Variation of the human mitochondrial genome (mtDNA) had been intensively studied and used to infer population history and human identification in forensic cases. Application of mtDNA profiling in Malaysia was however delayed due to lack of knowledge regarding the make up of the Malaysian Peninsular mtDNA pool. As a starting point, before any implementation of mtDNA analysis is possible in Malaysia, two populations of Peninsular Malaysia, the Modern Malays and Orang Asli, were chosen to be studied.

7.1 Sequencing of Hypervariable Region I

Sequence databases are the best source of information regarding the power of mtDNA for identity testing. Sequence data of mtDNA control region had been characterised from various populations worldwide. Direct sequencing of the amplified HVI region for Modern Malay and Orang Asli population had successfully identified 75 and 13 different haplotypes, respectively. Only five haplotypes were shared between the Modern Malay and Orang Asli, suggesting limited relation between both ancestors.

The existence of mtDNA as haploid molecules in each human being had made the statistical analysis for mtDNA sequence data much easier than any other genetic marker. Frequency of each mtDNA type was calculated by direct counting method. From the HVI sequence data, the Modern Malays showed high mtDNA variation than the Orang Asli. Higher mtDNA variation in the Modern Malays was expected since

their evolution was influenced by their ancestors trading and seafaring ways. The Modern Malays were believed to be the descendants of Proto Malays mixed with other races such as Indian, Chinese and Arab, whom they traded with. The present Modern Malays are scattered throughout Peninsular Malaysia, driven by their economy and social activities. The reduced variability observed in the Orang Asli might be a result of small sample size and sampling of maternally related individuals. Other possibility that might contribute to this phenomenon was the founder effect, which is an important cause of genetic drift.

Two point mutation heteroplamy were detected within the Orang Asli samples. The first heteroplasmic point mutation (at np 16,129) was detected by direct sequencing of the amplified HVI fragment and the second heteroplasmic point mutation (at np 16,243) was detected by sequencing of the cloned HVI fragment. This information is valuable as heteroplasmy was proven to improve probability of match (Ivanov *et al.*, 1996) and significantly increased the use of mtDNA as a forensic tool.

Heteroplasmy was not easily detected if the sequence quality was poor, with high background artefacts. It was therefore very important to keep high quality sequence when dealing with mtDNA analysis. The ABI Prism Big Dye® Terminator Cycle Sequencing Kit used for sequencing of the amplified HVI region had provided good quality of sequence for the purpose of this study.

7.2 Restriction Fragment Length Polymorphism Analysis of mtDNA Coding Region

RFLP analysis was the primary method being used to study the sequence variation of mtDNA coding region. This was undertaken to improve the classification of mtDNAs into haplogroups. RFLP analysis of the mtDNA coding region, along with the HVI sequence variations, had classified most of the Modern Malay samples into 3 major Southeast Asian specific haplogroups (M, B and F), suggesting a common lineage was shared between the Modern Malays and other Southeast Asian populations. The Modern Malays was also showing a small degree of affinity to the Indian population. This might be a result of intermarriage between both ancestors, especially during the famous Malacca Empire in the 13th century, which attracted many of the Indians sea-merchants migrated to Malay Peninsular.

The Orang Asli samples were categorized into two sub-clusters, M21a and R21 by using both RFLP and HVI sequence data. These two novel sub-clusters were also found in other Semang sub-groups studied before (Martin Richards, unpublished data). The Jahai showed high frequency of R21 while the Kinsiu showed higher frequency of M21a. The small number of mtDNA haplotypes found in both Semang sub-groups studied here might be a result of genetic drift. This can be supported by the small number of the Semang population, which appear to be the smallest group of Orang Asli in Peninsular Malaysia (less than 3000 individuals in 2002, Malaysian Department of Orang Asli Affairs). Other explanation to the low mtDNA variation in both Orang Asli sub-groups was the small sample size and sampling from maternally related individuals.

In forensic point of view, additional information gained from RFLP analysis of the mtDNA coding region did not add to higher power of discrimination when combined with the HVI sequence data. The definitive coding region mutations, which used for haplogroup determination, were always found in parallel with specific HVI polymorphisms. In cases where two individuals showed similar HVI sequence data, information from the definitive coding region mutations did not help to improve discrimination. Further analyses of the mtDNA coding region polymorphisms, especially for the non-definitive mutations should be perform in order to achieve better discrimination than using sequence data of the hypervariable regions alone.

Overall, the Orang Asli and Modern Malay populations are different in respect to their haplogroup composition. The Modern Malays, which exhibit high mtDNA diversity, did not appear to be direct descendants of the Orang Asli population.

7.3 Minisequencing Analysis

Human identification process based on mtDNA analysis had always been limited by the low power of discrimination. Single nucleotide polymorphisms (SNPs) of the mtDNA coding region were seen as the best candidates to be used along with hypervariable sequence data for higher discrimination. Since sequencing of the whole coding region is not very practical due to its larger size compared to hypervariable regions, a more efficient screening method is essential.

Minisequencing analysis of the mtDNA coding region mutations was performed using the SNaPshot™ Multiplex Kit. Two multiplex sets, which contained 10 and 7

minisequencing primers each, had successfully identified 17 SNPs that responsible for creating RFLP sites used in haplogroup determination. Results obtained from minisequencing analysis correspond to the RFLP and HVI sequence data gained earlier.

Minisequencing analysis using the SNaPshot™ Multiplex Kit was proven to be a robust and straightforward technique for screening of mtDNA coding region SNPs. In future, due to its relatively simple and fast technique, this method could replace routine sequencing of the hypervariable regions and also used for screening of other non-definitive coding region mutations. Haplogroups assignment in population studies will be much easier with application of this technique when compared to the classical restriction analysis, which is laborious and time consuming. The high sensitivity of this method was also an interesting feature for dealing with forensic samples, which often severely degraded or in minute amount. The ability for multiplexing had also reduced the time and effort when dealing with huge amount of samples. Since data interpretation was fully automated, it allows exact base identity and reduced the source of error.

7.4 Future work

i) Additional collection of the Modern Malay and Orang Asli samples in order to improve the statistical value and power of discrimination.

ii) Further analysis of hypervariable region II for both populations, which can provide more information and variations.

iii) Analysis of Y chromosome SNPs for both Modern Malay and Orang Asli populations would be an interesting area of future study in order to compare the population genetic structure of the paternal side to the result obtained by mtDNA analysis.

iv) Screening for non-definitive coding region polymorphisms, which can be used along with the hypervariable region sequence information for better power of discrimination.

CHAPTER 8: REFERENCES

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A, Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. and Young, I.G. (1981) Sequence and organisation of the human mitochondrial genome. *Nature*, **290**, 457-465.

Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M. and Howell. N. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature Genetics*, **23**, 147.

Aquadro, C.F. and Greenberg, B.D. (1983) Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. *Genetics*, **103**, 287-312.

Baasner, A., Schafer, C., Junge, A. and Madea, B. (1998) Polymorphic sites in human mitochondrial DNA control region sequences: population data and maternal inheritance. *Forensic Science International*, **98**, 169-178.

Ballinger, S.W., Schurr, T.G., Torroni, A., Gan, Y.Y., Hodge, J.A., Hassan, K., Chen, K.H. and Wallace, D.C. (1992) Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient Mongoloid migrations. *Genetics*, **130**, 139-152.

Bellwood, P. (1997) Prehistory of the Indo-Malaysian archipelago. University of Hawaii Press, Honolulu.

Bellwood, P. (1993) Cultural and biological differentiation in Peninsular Malaysia: the last 10,000 years. *Asian Perspectives*, **32**, 37-60.

Bellwood, P. (1985) Prehistory of the Indo-Malaysian Archipelago. Sydney, Academic Press.

Bendall, K.E and Sykes, B.C. (1995) Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. *American Journal of Human Genetics*, **57(2)**, 248-256.

Bender, K., Schneider, P.M. and Rittner, C. (2000) Application of mtDNA sequence analysis in forensic casework for the identification of human remains. *Forensic Science International*, **113**, 103-107.

Bogenhagen, D. and Clayton, D.A. (1974) The number of mitochondrial deoxyribonucleic acid genomes in mouse L and human HeLa cells. *Journal of Biological Chemistry*, **249**, 7991-7995.

Boles, T.C., Snow, C.C. and Stover, E. (1995) Forensic DNA testing on skeletal remains from mass graves: a pilot project in Guatemala. *Journal of Forensic Science*, **40(3)**, 349-355.

Brandstatter, A., Parsons, T.J. and Parson, W. (2003) Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups. *International Journal of Legal Medicine*, **117(5)**, 291-298.

Brown, M.D., Hosseini, S.H., Torroni, A., Bandelt, H.J., Allen, J.C., Schurr, T.G., Scozzari, R., Cruciani, F. and Wallace, D.C. (1998) mtDNA haplogroup X: An ancient link between Europe/Western Asia and North America? *American Journal of Human Genetics*, **63(6)**, 1852-1861.

Brown, W.M. (1980) Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. *Proceedings of the National Academy of Sciences of USA*, **77(6)**, 3605-9.

Budowle, B., Allard, M.W. and Wilson, M.R. (2002) Critique of interpretation of high levels of heteroplasmy in the human mitochondrial DNA hypervariable region I from hair. *Forensic Science International*, **126**, 30-33.

Budowle, B., DiZinno, J.A. and Wilson, M.R. (1999) Interpretation Guidelines for Mitochondrial DNA Sequencing, In: Proceeding of the Tenth International Symposium on Human Identification, Promega Corporation, Madison, WI. http://www. Promega.com/ussymp10proc/default.htm.

Budowle, B., Wilson, M.R., DiZinno, J.A., Stauffer, C., Fasano, M.A., Holland, M.M. and Monson, K.L. (1999b) Mitochondrial DNA regions HVI and HVII population data. *Forensic Science International*, **103**, 23-35.

Burckhardt, F., von-Haeseler, A. and Meyer, S. (1999) HvrBase: Compilation of mtDNA control region sequences from primates. *Nucleic Acids Research*, **27(1)**, 138-142.

Butler, J.M. (2001) Forensic DNA typing: Biology and Technology behind STR Markers. *Academic Press.*

Butler, J.M and Levin, B.C. (1998) Forensic applications of mitochondrial DNA. *Trends in Biotechnology*, **16(4)**, 158-62.

Cann, R.L., Stoneking, M. and Wilson, A.C. (1987) Mitochondrial DNA and human evolution. *Nature*, **325**, 31-36.

Carey, I. (1976) Orang Asli: the aboriginal tribes of Peninsular Malaysia. Oxford University Press, Kuala Lumpur.

Carracedo, A., Bar, W., Lincoln, P., Mayr, W., Morling, N., Olaisen, B., Schneider, P., Budowle, B., Brinkmann, B., Gill, P., Holland, M., Tully, G. and Wilson, M. (2000) DNA Comission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing. *Forensic Science International*, **110**, 79-85.

Chen, Y-S., Olckers, A., Schurr, T.G., Kogelnik, A.M., Huoponen, K. and Wallace, D.C. (2000) mtDNA variation in the South African Kung and Khwe-and their genetic relationships to other African populations. *American Journal of Human Genetics*, **66**, 1362-1383.

Ciminelli, B.M., Pompei, F., Malaspina, P., Hammer, M., Persichetti, F., Pignatti, P.F., Palena, A., et al. (1995) Recurrent simple tandem repeat mutations during human Y-chromosome radiation in Caucasian subpopulations. *Journal of Molecular Evolution*, **41**, 966-973.

Coble MD, Just RS, O'Callaghan JE, Letmanyi IH, Peterson CT, Irwin JA, Parsons TJ. (2004) Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians. *International Journal of Legal Medicine*, [Epub ahead of print].

Comas, D., Paabo, S. and Bertranpetit, J. (1995) Heteroplasmy in the control region of human mitochondrial DNA. *Genome Research*, **5**, 89-90.

Derbeneva, O.A., Starikovskaya, E.B., Wallace, D.C. and Sukernik, R.I. (2002) Traces of Early Eurasian in the Mansi of Northwest Siberia Revealed by Mitochondrial DNA Analysis. *American Journal of Human Genetics*, **70**, 1009-1014.

Denaro, M., Blanc, H., Johnson, M.J., Chen, K.H., Wilmsen, E., Cavalli-Sforza, L.L., Wallace, D.C. (1981) Ethnic variation in *Hpa*1 endonuclease cleavage patterns of human mitochondrial DNA. *Proceedings of the National Academy of Sciences of USA*, **78(9)**, 5768-72.

Edwards, A., Civitello, A., Hammond, H.A. and Caskey, C.T. (1991) DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *American Journal of Human Genetics*, **49(4)**, 746-756.

Endicott, P., Gilbert, T.M.P, Stringer, C., Lalueza-Fox, C., Willerslev, E., Hansen, A.J. and Cooper, A. (2003) The genetic origins of the Andaman islanders. *American Journal of Human Genetics*, **72**, 178-184.

Erdogan, F., Kirchner, R., Mann, W., Ropers, H.H. and Nuber, U.A. (2001) Detection of mitochondrial single nucleotide polymorphisms using a primer elongation reaction on oligonucleotide microarrays. *Nucleic Acids Research*, **29(7)**, E36.

Fiorentino, F., Magli, M.C., Podini, D., Ferraretti, A.P., Nuccitelli, A., Vitale, N., Baldi, M. and Gianaroli, L. (2003) The minisequencing method: an alternative strategy for preimplantation genetic diagnosis of single gene disorders. *Molecular Human Reproduction*, **9(7)**, 399-410.

Finnila, A., Lehtonen, M.S. and Majamaa, K. (2001) Phylogenetic network for European mtDNA. *American Journal of Human Genetics*, **68**, 1475-1474.

Fix, A.G. (1995) Malayan paleosociology: implications for patterns of genetic variation among the Orang Asli. *American Anthropology*, **97**, 313-323.

Forster, P., Harding, R., Torroni, A. and Bandelt, H.J. (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. *American Journal of Human Genetics*, **59(4)**, 935-945.

Foster, E.A., Jobling, M.A., Taylor, P.G., Donnelly, P., de Knijff, P., Mierement, R., Zerjal, T. and Tyler-Smith, C. (1998) Jefferson fathered slave's last child. *Nature*, **396**, 27-28.

Fucharoen, G., Fucharoen, S. and Horai, S. (2001) Mitochondrial DNA polymorphisms in Thailand. *Journal of Human Genetics*, **46**, 115-125.

Garritsen, H.S., Hoerning, A., Hellenkamp, F., Cassens, U., Mittmann, K., and Sibrowski, W. (2001) Polymorphisms in the non-coding region of the human mitochondrial genome in unrelated plateletapheresis donors. *British Journal of Haematology*, **112(4)**, 995-1003.

Gill, P., Urquhart, A., Millican, E., Oldroyd, N., Watson, S., Sparkes, R. and Kimpton, C.P. (1996) A new method of STR interpretation using inferential logic-development of a criminal intelligence database. *International Journal of Legal Medicine*, **109(1)**, 14-22.

Gill, P., Ivanov, P.L., Kimpton, C., Piercy, R., Benson, N., Tully, G., Evett, I., Hagelberg, E. and Sullivan, K. (1994) Identification of the remains of the Romanov family by DNA analysis. *Nature Genetics*, **6(2),** 130-135.

Ginther, C., Issel-Tarver, L. and King, M-C. (1992) Identifying individuals by sequencing mitochondrial DNA from teeth. *Nature Genetics,* **2(2),** 135-138.

Green, L.D., Derr, J.N. and Knight, A. (2000) mtDNA affinities of the peoples of north-central Mexico. *American Journal of Human Genetics*, **66**, 989-998.

Greenberg, B.D., Newbold, J.E. and Sugino, A. (1983) Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. *Gene*, **21**, 33-49.

Grzybowski, T. (2000) Extremely high levels of human mitochondrial DNA heteroplasmy in single hair roots. *Electrophoresis*, **21(3)**, 548-553.

Hadi, S.S. (2001) A molecular study of Pakistani populations using short tandem repeat markers. Department of Forensic Medicine and Science, University of Glasgow, p 56.

Hagelberg, E. and Clegg, J.B. (1993) Genetic polymorphisms in prehistoric Pacific islanders determined by analysis of ancient bone DNA. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **252(1334)**, 163-170.

Handt, O., Meyer, S., von Haeseler, A. (1998) Compilation of human mtDNA control region sequences. *Nucleic Acids Research*, **26(1)**, 126-129.

Harihara, S., Saitou, N., Hirai, M. *et al.* (1988) Mitochondrial DNA polymorphism among five Asian populations. *American Journal of Human Genetics*, **43**, 134-143.

Herrnstadt, C., Elson, J., Fahy, E., Preston, G., Turnbull, D.M., Anderson, C., Ghosh, S.S., Olefsky, J.M., Beal., M.F., Davis, R.E. and Howell, N. (2002) Reduced-Median-Network Analysis of Complete Mitochondrial DNA Coding-Region Sequences

for the Major African, Asian and European Haplogroups. *American Journal of Human Genetics*, **70**, 1152-1171.

Hertzberg, M., Mickleson, K.N., Serjeantson, S.W., Prior, J.F. and Trent, R.J. (1989) An Asian-specific 9-bp deletion of mitochondrial DNA is frequently found in Polynesians. *American Journal of Human Genetics*, **44(4)**, 504-510.

Holland, M.M. and Parsons, T.J. (1999) Mitochondrial DNA Sequence Analysis-Validation and Use for Forensic Casework. *Forensic Science Review*, **11**, 21-50.

Hopwood, A.J., Mannucci, A. and Sullivan, K.M. (1996) DNA typing from human faeces. *International Journal of Legal Medicine*, **108(5)**, 237-243.

Hopgood, R., Sullivan, K.M. and Gill, P. (1992) Strategies for automated sequencing of human mitochondrial DNA directly from PCR products. *BioTechniques,* **13**, 82-92.

Horai, S., Murayama, K., Hayasaka, K., Matsubayashi, S., Hattori, Y., Fucharoen, G., Harihara, S., Park, K.S., Omoto, K. and Pan, I.H. (1996) mtDNA polymorphism in East Asian Populations, with special reference to the peopling of Japan. *American Journal of Human Genetics*, **59(3)**, 579-590.

Horai, S., Kondo, R., Nakagawa-Hattori, Y., Hayashi, S., Sonoda, S. and Tajima, K. (1993) Peopling of the Americas, founded by four major lineages of mitochondrial DNA. *Molecular Biology and Evolution*, **10(1)**, 23-47.

Horai, S., Gojobori, T. and Matsunaga, E. (1984) Mitochondrial DNA polymorphism in Japanese. Analysis with restriction enzymes of six base pair recognition. *Human Genetics*, **68**, 324-332.

Imaizumi, K., Parsons, T.J., Yoshino, M. and Holland, M.M. (2002) A new database of mitochondrial DNA hypervariable regions I and II sequences from 162 Japanese individuals. *International Journal of Legal Medicine*, **116**, 68-73.

Ivanov, P.L., Wadhams, M.J., Roby, R.K., Holland, M.M., Weedn, V.W. and Parsons, T.J. (1996) Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. *Nature Genetics*, **12(4)**, 417-420.

Jeffreys, A.J., Wilson, V. and Thein,S.L. (1985) Individual specific fingerprints of human DNA. *Nature*, **316**, 76-79.

Jobling, M.A. and Tyler-Smith, C. (2000) New uses for new haplotypes the human Y chromosome, disease and selection. *Trends in Genetics*, **16(8)**, 356-362.

Jobling, M.A. and Tyler-Smith, C. (1995) Fathers and sons: the Y chromosome and human evolution. *Trends in Genetics*, **11**, 449-456.

Kimpton, C., Fisher, D., Watson, S., Adams, M., Urquhart, A., Lygo, J. and Gill, P. (1994) Evaluation of an automated DNA profiling system employing multiplex amplification of four tetrameric STR loci. *International Journal of Legal Medicine*, **106(6)**, 302-311.

Kimpton, C.P., Gill, P., Walton, A., Urquhart, A., Millican, E.S. and Adams, M. (1993) Automated DNA profiling employing multiplex amplification of short tandem repeat loci. *PCR Methods and Applications*, **3(1)**, 13-22.

Kivisild,T., Tolk, H-V., Parik, J., Wang, Y., Papiha, S.S., Bandelt, H-J. and Villems, R. (2002) The emerging limbs and twigs of the East Asian mtDNA tree. *Molecular Biology and Evolution*, **19(10)**, 1737-1751.

Kivisild, T., Kaldma, K., Metspalu, M., Parik, J., Papiha, S. and Villems, R. (1999) The place of the Indian mitochondrial DNA variants in the global network of maternal lineages and the peopling of the old world. *Genomic Diversity: Applications in Human Population Genetics*. Kluwer Academic/Plenum Publisher, New York 135-152.

Kocher, T.D. and Wilson, A.C. (1991) Sequence evolution of mitochondrial DNA in humans and chimpanzees: control region and protein-coding regions. In *Evolution of*

Life: Fossils, Molecules and Cultures, edited by Osawa, S. and Honio, T. Springer Verlag, Tokyo. pp 391-413.

Kurosaki, K., Matsushita, T. and Ueda, S. (1993) Individual DNA identification from ancient human remains. *American Journal of Human Genetics*, **53(3)**, 638-643.

Landegren, U., Kaiser, R., Sanders, J. and Hood, L. (1988) A ligase-mediated gene detection technique. *Science*, **24**, 1077-1080.

Lagerstrom-Fermer, M., Olsson, C., Forsgren, L. and Syvanen, A-C. (2001) Heteroplasmy of the human mtDNA control region remains constant during life. *American Journal of Human Genetics*, **68**, 1299-1301.

Lee, L.G., Connell, C.R. and Bloch, W. (1993) Allelic discrimination by nick-translation PCR with fluorogenic probes. *Nucleic Acid Research*, **21**, 3761-3766.

Lee, S.D., Shin, C.H., Kim, K.B., Lee, Y.S. and Lee, J.B. (1997) Sequence variation of mitochondrial DNA control region in Koreans. *Forensic Science International*, **87(2)**, 99-116.

Lum, J.K., Rickards, O., Ching, C. and Cann, R.L. (1994) Polynesian mitochondrial DNAs reveal three deep maternal lineage clusters. *Human Biology*, **66(4)**,567-590.

Lutz, S., Weisser, H-J., Heizmann, J. and Pollak, S. (1997) A third hypervariable region in the human mitochondrial D-loop. *Human Genetics*, **101(3)**, 384.

Lyamichev, V., Mast, A.L., Hall, J.G., Prudent, J.R., Kaiser, M.W., Takova, T., Kwiatkowski, R.W., Sander, T.J., deArruda, M., Arco, D.A., Neri, B.P. and Brow, M.A.D. (1999) Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nature Biotechnology*, **17**, 292-296.

Macaulay, V., Richards, M., Hickey, E., Vega, E., Cruciani, F., Guida, V., Scozzari, R., Bonne-Tamir, B., Sykes, B. and Torroni, A. (1999) The emerging tree of West

Eurasian mtDNAs: A synthesis of control-region sequences and RFLPs. *American Journal of Human Genetics*, **64**, 232-249.

Melton, T. and Nelson, K. (2001) Forensic Mitochondrial DNA analysis: Two years of commercial casework experience in the United States. *Croatian Medical Journal*, **42(3)**, 298-303.

Melton, T., Peterson, R., Redd, A.J., Saha, N., Sofro, A.S., Martinson, J. and Stoneking, M. (1995) Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. *American Journal of Human Genetics*, **57(2)**, 403-414.

Meyer, S., Weiss, G. and von-Haeseler, A. (1999) Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. *Genetics*, **152**, 1103-1110.

Miller, K.W. and Budowle, B. (2001) A compendium of human mitochondrial DNA control region: development of an international standard forensic database. *Croatian Medical Journal*, **42(3)**, 315-327.

Miller, K.W., Dawson, J.L. and Hagelberg, E. (1996) A concordance of nucleotide substitutions in the first and second hypervariable segments of the human mtDNA control region. *International Journal of Legal Medicine*, **109(3)**,107-113.

Morley, J.M., Bark, J.E., Evans, C.E., Perry, J.G., Hewitt, C.A. and Tully, G. (1999) Validation of mitochondrial DNA minisequencing for forensic casework. *International Journal of Legal Medicine*, **112**, 241-248.

Mountain, J.L., Hebert, J.M., Bhattacharyya, S., Underhill, P.A., Ottolenghi, C., Gadgil, M. and Cavalli-Sforza, L.L. (1995) Demographic history of India and mtDNA sequence diversity. *American Journal of Human Genetics*, **56**, 979-992.

Mullis, K.B., Faloona, F., Scharf, S.J., Saiki, R.K., Horn, G.T. and Erlich, H.A. (1986) Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. *Cold Spring Harbor Symposium on Quantitative Biology*, **51**, 263-273.

Nishimaki, Y., Sato, K., Fang, L., Ma, M., Hasekura, H. and Boettcher, B. (1999) Sequence polymorphism in the mtDNA HV1 region in Japanese and Chinese. *Legal Medicine (Tokyo)*, **1(4)**, 238-249.

Paabo, S. (1999) Human evolution. *Trends in Cell Biology*, **9(12)**, 13-16.

Parsons, T.J. and Coble, M.D. (2001) Increasing the forensic discrimination of mitochondrial DNA testing through analysis of the entire mitochondrial genome. *Croatian Medical Journal*, **42(3)**, 304-309.

Parson, W., Parsons, T.J., Scheithauer, R. And Holland, M.M. (1998) Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: application of mtDNA sequence analysis to a forensic case. *International Journal of Legal Medicine*, **111(3)**, 124-132.

Penta, J.S., Johnson, F.M., Wachsman, J.T. and Copeland, W.C. (2001) Mitochondrial DNA in human malignancy. *Mutation Research*, **488**, 119-133.

Piercy, R., Sullivan, K.M., Benson, N. and Gill, P. (1993) The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. *International Journal of Legal Medicine*, **106(2)**, 85-90.

Quintana-Murci, L., Semino, O., Bandelt, H.J., Passarino, G., McElreavey, K. and Santachiara-Benerecetti, A.S. (1999) Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa. *Nature Genetics*, **23(4)**, 437-441.

Rando, J.C., Pinto, F., Gonzalez, A.M., Hernandez, M., Larruga, J.M., Cabrera, V.M. and Bandelt, H.J. (1998) Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations. *Annals of Human Genetics*, **62**, 531-550.

Rasmussen, E.M., Sorensen, E., Eriksen, B., Larsen, H.J. and Morling, N. (2002) Sequencing strategy of mitochondrial HV1 and HV2 DNA with length heteroplasmy. *Forensic Science International*, **129(3)**, 209-213.

Rayner, D. and Bulbeck, D. (2001) Dental morphology of the 'Orang Asli' aborigines of the Peninsular Malaysia. In: Henneberg, M. (ed) *Causes and effects of human variation.* Australasian Society for Human Biology, University of Adelaide, pp 19-41.

Redd, A.J., Takezaki, N., Sherry, S.T., McGarvey, S.T., Sofro, A.S. and Stoneking M. (1995) Evolutionary history of the COII/tRNALys intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. *Molecular Biology and Evolution*, **12(4)**, 604-615.

Richards, B., Skoletsky, J., Shuber, A.P., Balfour, R., Stern, R.C., Dorkin, H.L., Parad, R.B., Witt, D. and Klinger, K.W. (1993) Multiplex PCR amplification from the CFTR gene using DNA prepared from buccal brushes/swabs. *Human Molecular Genetics*, **2(2)**, 159-163.

Richards, M., Rengo, C., Cruciani, F., Gratrix, F., Wilson, J.F., Scozzari, R., Macaulay, V. and Torroni, A. (2003) Extensive female-mediated gene flow from sub-Saharan Africa into near eastern Arab populations. *American Journal of Human Genetics*, **72(4)**, 1058-64.

Richards, M. and Macaulay, V. (2000) Genetic data and the colonization of Europe: Genealogies and founders. In *Archaeogenetics: DNA and the population prehistory of Europe*, Edited by Renfrew, C. and Boyle, K. McDonald Institute for Archaeological Research, Cambridge, UK. pp 139-151.

Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Guida, V., Rengo, C., Sellitto, D., Cruciani, F., Kivisild, T., Villems, R., Thomas, M., Rychkov, S., Rychkov, O., Rychkov, Y. *et al.* (2000) Tracing European Founder Lineages in the Near Eastern mtDNA Pool. *American Journal of Human Genetics*, **67**, 1251-1276.

Rudin, N. and Inman, K. (2002) An Introduction to Forensic DNA Analysis, 2nd Edition, CRC Press, New York.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239**, 487-491.

Schneider, P.M. (1997) Basic issue in forensic DNA typing. *Forensic Science International*, **88**, 17-22.

Schurr, T.G. and Wallace, D.C. (2002) Mitochondrial DNA diversity in Southeast Asian populations. *Human Biology*, **74(3)**, 431-452.

Schurr, T.G., Ballinger, S.W., Gan, Y.Y., Hodge, J.A., Merriwether, D.A., Lawrence, D.N., Knowler, W.C., Weiss, K.M. and Wallace, D.C. (1990) Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. *American Journal of Human Genetics*, **46(3)**, 613-623.

Scozzari, R., Cruciani, F., Malaspina, P., Santolamazza, P., Ciminelli, B.M., Torroni, A., Modiano, D., Wallace, D.C., Kidd, K.K., Olckers, A., Moral, P., Terrenato, L., Akar, N., Qamar, R., Mansoor, A., Mehdi, S.Q., Meloni, G., Vona, G., Cole, D.E., Cai, W. and Novelletto, A. (1997) Differential structuring of human populations for homologous X and Y microsatellite loci. *American Journal of Human Genetics*, **61(3)**, 719-733.

Seo, Y., Stradmann-Bellinghausen, B., Rittner, C., Takahama, K. and Schneider, P.M. (1998) Sequence polymorphism of mitochondrial DNA control region in Japanese. *Forensic Science International*, **97**, 155-164.

Shields, G.F., Hecker, K., Voevoda, M.I. and Reed, J.K. (1992) Absence of the Asian-specific region V mitochondrial marker in Native Beringians. *American Journal of Human Genetics*, **50(4)**, 758-765.

Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Kent, S.B. and Hood, L.E. (1986) Fluorescence detection in automated DNA sequence analysis. *Nature*, **321**, 674-679.

Sparkes, R., Kimpton, C., Gilbard, S., Carne, P., Andersen, J., Oldroyd, N., Thomas, D., Urquhart, A. and Gill, P. (1996) The validation of a 7-locus multiplex STR test for use in forensic casework. (II), Artefacts, casework studies and success rates. *International Journal of Legal Medicine*, **109(4)**, 195-204.

Stewart, J.E.B., Fisher, C.L., Aagaard, P.J., Wilson, M.R., Isenberg, A.R., Polanskey, D., Pokorak, E., DiZinno, J.A. and Budowle, B. (2001) Length variation in HV2 of the human mitochondrial DNA control region. *Journal of Forensic Science*, **4**, 862-870.

Stoneking, M., Hedgecock, D., Highchi, R.G., Vigilant, L. and Erlich, H.A. (1991) Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. *American Journal of Human Genetics*, **48**, 370-382.

Stoneking, M., Jorde, L.B., Bhatia, K. and Wilson, A.C. (1990) Geographic variation in human mitochondrial DNA from Papua New Guinea. *Genetics*, **124(3)**, 717-733.

Sullivan, K.M., Alliston-Greiner, R., Archampong, F.I.A., Piercy, R., Tully, G., Gill, P. and Lloyd-Davies, C. (1996) A Single Difference in mtDNA Control Region Sequence Observed Between Hair Shaft and Reference Samples from a Single Donor. *Proceedings from the Seventh International Symposium on Human Identification* pp 126-130, Promega Corporation.

Sullivan, K.M., Hopgood, R. and Gill, P. (1992) Identification of human remains by amplification and automated sequencing of mitochondrial DNA. *International Journal of Legal Medicine*, **105**, 83-86.

Sullivan, K.M., Hopgood, R., Lang, B. and Gill, P. (1991) Automated amplification and sequencing of human mitochondrial DNA. *Electrophoresis*, **12(1)**, 17-21.

Syvanen, A., Aalto-Setala, K., Kontula, K. and Sonderlund, H. (1990) A primer-guided nucleotide incorporation assay in genotyping of apolipoprotein E. *Genomics*, **8**, 684-692.

Szibor, R., Michael, M., Plate, I. and Krause, D. (2000) Efficiency of forensic mtDNA analysis case examples demonstrating the identification of traces. *Forensic Science International*, **113**, 71-78.

Taanman, J-W. (1999) The mitochondrial genome, transcription, translation and replication. *Biochimica et Biophysica Acta*, **1410**, 103-123.

Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123(3)**, 585-595.

Tamura, K. and Nei, M. (1993) Estimation of the number of nucleotide substitution in the control region of mitochondrial DNA in humans and chimpanzee. *Molecular Biology and* Evolution, **10**, 512-526.

Taylor, P.L. (1990) GeneJockey Sequence Processor Software, Biosoft, Cambridge, UK.

Torroni, A., Bandelt, H.J., D'Urbano, L., Lahermo, P., Moral, P., Sellitto, D., Rengo, C., Forster, P., Savontaus, M.L., Bonne-Tamir, B. and Scozzari, R. (1998) mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *American Journal of Human Genetics*, **62(5)**, 1137-52.

Torroni, A., Huoponen, K., Francalacci, P., Petrozzi, M., Morelli, L., Scozzari, R., Obinu, D., Sarontaus, M-L. and Wallace, D.C. (1996) Classification of European mtDNAs from an analysis of three European populations. *Genetics*, **144**, 1835-1850.

Torroni, A. and Wallace, D.C. (1994) Mitochondrial DNA variation in human populations and implications for detection of mitochondrial DNA mutations of pathological significance. *Journal of Bioenergetics and Biomembranes*, **26**, 261-271.

Torroni, A., Miller, J.A., Moore, L.G., Zamudio, S., Zhuang, J., Droma, T. and Wallace, D.C. (1994) Mitochondrial DNA analysis in Tibet: Implications for the origin

of the Tibetan population and its adaptation to high altitude. *American Journal of Physical Anthropology*, **93(2)**, 189-199.

Torroni, A., Schurr, T.G., Cabell, M.F., Brown, M.D., Neel, J.V., Larsen, M., Smith, D.G., Vullo, C.M. and Wallace, D.C. (1993) Asian affinities and continental radiation of the four founding Native American mtDNAs. *American Journal of Human Genetics*, **53**, 563-590.

Torroni, A., Schurr, T.G., Yang, C-C., Szathmary, E.J.E., Williams, R.C., Schanfield, M.S., Troup, G.A., Knowler, W.C., Lawrence, D.N., Weiss, K.M. and Wallace, D.C. (1992) Native American mitochondrial DNA analysis indicates that the Amerind and Nadene populations were founded by two independent migrations. *Genetics*, **130**, 153-162.

Tsai, L.C., Lin, C.Y., Lee, J.C.I., Chang, J.G., Linacre, A. and Goodwin, W. (2001) Sequence polymorphisms of mitochondrial D-loop DNA in the Taiwanese Han population. *Forensic Science International*, **119**, 239-247.

Tully, G., Sullivan, K.M., Nixon, P., Stones, R.E. and Gill, P. (1996) Rapid detection of mitochondrial sequence polymorphisms using multiplex solid-phase fluorescent minisequencing. *Genomics*, **34**, 107-113.

Upholt, W.B. and David, I.B. (1977) Mapping of mtDNA of individual sheep and goat: rapid evolution in the D-loop region. *Cell*, **11**, 571-583.

Urquhart, A., Oldroyd, N.J., Kimpton, C.P. and Gill, P. (1995) Highly discriminating heptaplex short tandem repeat PCR system for forensic identification. *Biotechniques*, **18(1)**, 116-8, 120-1.

Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. and Wilson, A.C. (1991) African populations and evolution of mitochondrial DNA. *Science*, **253**, 1503-1507.

Wakeley, J. (1993) Substitution rate variation among sites in hypervariable region I of human mitochondrial DNA. *Journal of Molecular Evolution*, **37**, 613-623.

Wallace, D.C. (1997) Mitochondrial DNA in ageing and disease. *Scientific American*, **277**, 40-47.

Wallace, D.C. (1995) Mitochondrial DNA variation in human evolution, degenerative disease and aging. *American Journal of Human Genetics*, **57**, 201-223.

Wallace, D.C. (1994) Mitochondrial DNA sequence variation in human evolution and disease. *Proceedings of the National Academy of Sciences of USA*, **91(19)**, 8739-8746.

Wallace, D.C, Lott, M.T., Brown, M.D., Huoponen, K. and Torroni, A. (1995) Report of the committee on human mitochondrial DNA. In: *Human Gene Mapping 1995: a Compendium*, edited by A.J.Cuticchia, p 910-954. Johns Hopkins University Press, Baltimore.

Walsh, P.S., Metzger, D.A. and Higuchi, R. (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques*, **10(4)**, 506-513.

Wang, D.G., Fan, J.B., Siao, C.J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Krug-lyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M.S., Shen, N.P., Kilburn, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T.J., Lipshutz, R., Chee, M. and Lander, E.S. (1998) Large-scale identification, mapping and genotyping of single nucleotide polymorphisms in the human genome. *Science*, **280**, 1077-1082.

Watson, E., Forster, P., Richards, M and Bandelt, H-J. (1997) Mitochondrial footprints of human expansions in Africa. *American Journal of Human Genetics*, **61**, 691-704.

Weber, J.L. and May, P.E. (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics*, **44(3)**, 388-396.

Wilson, A.C., Cann, R.L., Carr, S.M., George, M., Gyllensten, G.B., Helm-Bychowski, K.M., Higuchi, R.G., Palumbi, S., Prager, E.M., Sage, R.D. and Stoneking, M. (1985) Mitochondrial DNA and two perspectives on evolutionary genetics. *Biological Journal of the Linnean Society*, **26**: 375-400

Wilson, M.R, DiZinno, J.A., Polanskey, D., Replogle, J. and Budowle, B. (1995a) Validation of mitochondrial DNA sequencing for forensic casework analysis. *International Journal of Legal Medicine*, **108**, 68-74.

Wilson, M.R., Polanskey, D., Butler, J.M., DiZinno, J.A., Replogle, J. and Budowle, B. (1995b) Extraction, PCR amplification and sequencing of mitochondrial DNA from human hair shafts. *BioTechniques*, **18(4)**, 662-669.

Wilson, M.R., Stoneking, M., Holland, M.M, DiZinno, J.A. and Budowle, B. (1993) Guidelines for the use of mitochondrial DNA sequencing in forensic science. *Crime Laboratory Digest*, **20**, 68-77.

Wittig, H., Augustin, C., Baasner, A., Bulnheim, U., Dimo-Simonin, N., Edelmann, J., Hering, S., Jung, S., Lutz, S., Michael, M., Parson, W., Poetsch, M., Schneider, P.M., Weichhold, G. and Krause, D. (2000) Mitochondrial DNA in the central European population: Human identification with the help of the forensic mtDNA D-loop-Base Database. *Forensic Science International*, **113**, 113-118.

Wei, Y.-H. (1992) Mitochondrial DNA alterations as ageing-associated molecular events. *Mutation Research*, **275**, 145-155.

Wong, Z., Wilson, V., Patel, I., Povey, S. and Jeffreys, A.J. (1987) Characterization of a panel of highly variable minisatellites cloned from human DNA. *Annals of Human Genetics*, **51**, 269-288.

Yao, Y.G., Kong, Q.P., Bandelt, H.J., Kivisild, T. and Zhang, Y.P. (2002) Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *American Journal of Human Genetics*, **70(3)**, 635-51.

APPENDIX

Genotype frequency for locus D21S11, FGA and PENTA E for Modern Malay and Orang Asli population in Peninsular Malaysia.

Locus: D21S11					Locus: FGA					Locus: PENTA E				
Genotype	MM		OA		Genotype	MM		OA		Genotype	MM		OA	
	P	SD	P	SD		P	SD	P	SD		P	SD	P	SD
29,28	0.031	0.034	0.055	0.060	19,19	0.010	0.019	0.018	0.035	7,7	0.010	0.019		
29,29	0.051	0.044	0.109	0.082	20,19			0.070	0.066	11,5	0.029	0.033		
30,28	0.031	0.034	0.018	0.035	20,20			0.018	0.035	11,7	0.010	0.019		
30,29	**0.184**	0.077	0.055	0.060	20.2,20	0.010	0.019			11,9	0.010	0.019		
30,30	0.010	0.019	0.036	0.049	21,19	0.010	0.019	0.018	0.035	11,10			0.018	0.035
30.2,28	0.020	0.028			21,20	0.029	0.033	0.035	0.048	11,11	**0.067**	0.049	**0.175**	0.099
30.2,29	0.020	0.028	0.091	0.076	21,21	0.020	0.027	0.018	0.035	12,5	0.010	0.019		
30.2,30	0.020	0.028	0.018	0.035	21.2,19	0.010	0.019	0.035	0.048	12,9	0.010	0.019		
30.2,30.2			0.036	0.049	21.2,20			0.018	0.035	12,11	0.010	0.019		
31,29	0.061	0.047			21.2,21	0.029	0.033	0.053	0.058	12,12	0.029	0.033		
31,30	0.051	0.044			22,19	0.059	0.046	0.053	0.058	13,10	0.010	0.019		
31,30.2	0.010	0.019			22,20	0.029	0.033	0.035	0.048	13,11	0.059	0.046	0.018	0.035
31,31	0.041	0.039			22,21	0.059	0.046	**0.105**	0.080	13,12	0.029	0.033		
31.2,29	0.020	0.028	**0.164**	0.098	22,22	0.069	0.049	0.035	0.048	13,13			0.018	0.035
31.2,30	0.082	0.054	0.055	0.060	22.2,20			0.018	0.035	14,9	0.029	0.033		
31.2,30.2	0.010	0.019	0.036	0.049	22.2,21	0.020	0.027	0.088	0.074	14,10			0.035	0.048
31.2,31	0.020	0.028	0.018	0.035	22.2,22	0.010	0.019	0.018	0.035	14,11	0.039	0.038	**0.175**	0.099
31.2,31.2	0.010	0.019	0.018	0.035	23,19	0.020	0.027			14,12	0.020	0.027		
32,31	0.020	0.028			23,20	0.020	0.027	0.053	0.058	14,13	0.010	0.019	0.017	0.034
32,32	0.010	0.019			23,20.2	0.010	0.019			14,14			0.070	0.066
32.2,28	0.041	0.039	0.036	0.049	23,21	0.059	0.046	0.053	0.058	15,5	0.029	0.033		
32.2,29	0.020	0.028	0.055	0.060	23,22	0.069	0.049	0.035	0.048	15,9	0.010	0.019		
32.2,30	0.071	0.051			23,22.2	0.010	0.019	0.018	0.035	15,11	0.029	0.033	0.070	0.066
32.2,31	0.010	0.019			23,23	0.020	0.027			15,12	0.029	0.033		
32.2,31.2	0.020	0.028	0.036	0.049	24,18	0.010	0.019			15,14	0.010	0.019		
32.2,32	0.010	0.019	0.018	0.035	24,19	0.010	0.019			16,11	0.039	0.038		
32.2,32.2	0.020	0.028			24,21	0.029	0.033	0.035	0.048	16,12	0.020	0.027		
33,32.2	0.010	0.019			24,22	**0.078**	0.052	0.035	0.048	16,13	0.010	0.019		
33.2,29	0.020	0.028	0.073	0.069	24,22.2	0.010	0.019	0.018	0.035	16,14	0.020	0.027		
33.2,30.2			0.018	0.035	24,23	0.029	0.033			17,11	0.029	0.033	0.070	0.066
33.2,31	0.041	0.039			24,24			0.018	0.035	17,12	0.010	0.019		
33.2,31.2			0.018	0.035	24.2,22.2					17,13	0.010	0.019		
33.2,32	0.010	0.019			24.2,23	0.010	0.019			17,14			0.035	0.048
33.2,32.2	0.010	0.019	0.036	0.049	25,19	0.020	0.027			17,15	0.010	0.019		
34.2,29	0.010	0.019			25,20	0.010	0.019	0.018	0.035	17,16	0.010	0.019		
					25,21			0.035	0.048	18,5	0.010	0.019		
					25,22	0.049	0.042			18,11	0.010	0.019		
					25,22.2	0.010	0.019			18,12	0.010	0.019		
					25,23	0.020	0.027	0.035	0.048	18,13	0.010	0.019	0.035	0.048
					25,24	0.029	0.033	0.018	0.035	18,14	0.010	0.019	0.035	0.048
					25.2,20	0.010	0.019			18,16	0.010	0.019		
					26,21	0.010	0.019			18,17	0.010	0.019		
					26,22	0.029	0.033			18,18	0.010	0.019	0.018	0.035
					26,23	0.010	0.019			19,11			0.070	0.066
					26,24	0.020	0.027			19,12	0.020	0.027		
					26.2,21	0.010	0.019			19,13	0.010	0.019		
					27,26	0.010	0.019			19,15			0.018	0.035
					28,20	0.010	0.019			19,16	0.010	0.019		
										19,18	0.020	0.027		
										20,11	0.010	0.019	0.035	0.048
										20,12	0.029	0.033		
										20,13	0.010	0.019		
										20,18	0.010	0.019	0.018	0.035
										20,19	0.010	0.019	0.018	0.035
										20,20				
										21,12	0.010	0.019		
										21,13	0.010	0.019		
										21,14	0.010	0.019		
										22,11	0.010	0.019		
										22,13	0.010	0.019		
										23,11			0.018	0.035
										23,14			0.018	0.035
										24,17	0.010	0.019		
										28,18			0.018	0.035

Genotype frequency for locus D18S51, Penta D and vWA for Modern Malay and Orang Asli population in Peninsular Malaysia.

Locus: D18S51					Locus: Penta D					Locus: vWA				
Genotype	MM		OA		Genotype	MM		OA		Genotype	MM		OA	
	P	SD	P	SD		P	SD	P	SD		P	SD	P	SD
12,10	0.010	0.019			9,8	0.020	0.027			14,14	0.039	0.038	0.070	0.066
12,10.2			0.018	0.033	9,9	**0.157**	0.071	0.053	0.058	15,14	0.020	0.027		
12,12	0.010	0.019			10,7	0.010	0.019			16,14	0.059	0.046	0.018	0.035
13,12	0.020	0.027			10,8	0.010	0.019			16,15	0.010	0.019		
13,13	0.010	0.019			10,9	**0.157**	0.071	0.088	0.074	16,16	0.039	0.038	0.018	0.035
14,12	0.039	0.038			10,10	0.020	0.027	0.035	0.048	17,14	0.069	0.049	0.123	0.085
14,13	0.029	0.033			11,7	0.020	0.027			17,15	0.010	0.019		
14,14	0.020	0.027	0.018	0.033	11,9	0.088	0.055	**0.228**	0.109	17,16	0.078	0.052	0.140	0.090
15,12	0.039	0.038	0.105	0.080	11,10	0.049	0.042	0.105	0.080	17,17	0.049	0.042	**0.211**	0.106
15,13	0.069	0.049	0.070	0.066	11,11	0.020	0.027	0.123	0.085	18,14	**0.147**	0.069	0.070	0.066
15,14	**0.118**	0.063	0.123	0.085	12,7	0.010	0.019			18,15	0.029	0.033		
15,15	0.039	0.038	**0.193**	0.102	12,8	0.029	0.033	0.018	0.035	18,16	0.029	0.033	0.070	0.066
16,10			0.018	0.033	12,9	0.069	0.049	0.035	0.048	18,17	0.127	0.064	0.140	0.090
16,12	0.010	0.019	0.018	0.033	12,10	0.059	0.046	0.018	0.035	18,18	0.049	0.042	0.010	0.026
16,13	0.039	0.038			12,11	0.039	0.038			19,14	0.039	0.038	0.010	0.026
16,14	0.039	0.038	0.053	0.058	12,12	0.029	0.033			19,16	0.010	0.019	0.010	0.026
16,14.2			0.018	0.033	13,9	0.069	0.049	0.035	0.048	19,17	0.088	0.055	0.070	0.066
16,15	0.078	0.052	**0.193**	0.102	13,10	0.049	0.041			19,18	0.029	0.033	0.018	0.035
16,16	0.029	0.033	0.088	0.074	13,11	0.020	0.027	0.018	0.035	20,14	0.010	0.019		
17,11	0.010	0.019			13,12	0.020	0.027	0.018	0.035	20,16	0.020	0.027		
17,12	0.029	0.033			13,13			0.018	0.035	20,17	0.020	0.027		
17,13	0.020	0.027			14,9	0.049	0.042	0.035	0.048	20,18	0.020	0.027		
17,14	0.020	0.027			14,10			0.018	0.035	20,19	0.010	0.019		
17,15	0.039	0.038			14,11	0.010	0.019	0.070	0.066					
17,16	0.039	0.038			15,9			0.018	0.035					
18,14	0.010	0.019	0.018	0.033	15,11			0.053	0.058					
18,15	0.049	0.042			15,13			0.018	0.035					
18,16	0.029	0.033												
19,14	0.010	0.019												
19,15	0.029	0.033	0.018	0.033										
19,16	0.020	0.027												
19,18	0.010	0.019												
20,12	0.020	0.027												
20,14	0.010	0.019												
20,17	0.010	0.019												
21,12	0.010	0.019												
21,15			0.053	0.058										
21,17	0.010	0.019												
23,12	0.010	0.019												
23,18	0.010	0.019												
24,17	0.010	0.019												

Genotype frequency for locus D7S820, D16S539 and TPOX for Modern Malay and Orang Asli population in Peninsular Malaysia.

Locus: D7S820					Locus: D16S539					Locus: TPOX				
Genotype	MM		OA		Genotype	MM		OA		Genotype	MM		OA	
	P	SD	P	SD		P	SD	P	SD		P	SD	P	SD
8,8	0.050	0.043	0.019	0.036	9,8			0.018	0.035	8,8	**0.333**	0.091	0.228	0.109
9,8	0.020	0.027			9,9	0.059	0.046	0.018	0.035	9,8	0.196	0.077	0.053	0.058
10,7	0.010	0.019			10,8	0.010	0.019			9,9	0.010	0.019	0.018	0.035
10,8	0.069	0.049	0.074	0.070	10,9	0.059	0.046	0.053	0.058	10,8	0.029	0.033	0.035	0.048
10,9	0.020	0.027			10,10	0.029	0.033	0.053	0.058	10,9	0.010	0.019	0.018	0.035
10,10	0.020	0.027	0.019	0.036	11,8	0.010	0.019	0.018	0.035	11,8	0.245	0.083	0.298	0.118
11,7	0.020	0.027			11,9	0.088	0.055	0.105	0.080	11,9	0.049	0.042	0.053	0.058
11,8	0.099	0.058	0.222	0.111	11,10	0.069	0.049	**0.175**	0.099	11,10	0.020	0.027	0.053	0.058
11,9	0.040	0.038			11,11	0.088	0.055	0.070	0.066	11,11	0.069	0.049	**0.246**	0.112
11,10	**0.198**	0.078	0.074	0.070	12,8			0.035	0.048	12,8	0.010	0.019		
11,11	0.069	0.049	**0.241**	0.114	12,9	**0.118**	0.063	0.088	0.074	12,9	0.010	0.019		
12,7	0.010	0.019			12,10	0.039	0.038	0.053	0.058	12,11	0.020	0.027		
12,8	0.079	0.053	0.056	0.061	12,11	0.098	0.058	0.088	0.074					
12,10	0.109	0.061	0.111	0.084	12,12	0.049	0.042	0.035	0.048					
12,11	0.119	0.063	0.111	0.084	13,8			0.035	0.048					
12,12	0.020	0.027	0.056	0.061	13,9	0.039	0.038							
13,8	0.010	0.019	0.019	0.036	13,10	0.020	0.027	0.053	0.058					
13,11	0.040	0.038			13,11	**0.118**	0.063	0.070	0.066					
					13,12	0.029	0.033	0.035	0.048					
					13,13	0.020	0.027							
					14,9	0.020	0.027							
					14,10	0.020	0.027							
					14,11	0.010	0.019							
					15,12	0.010	0.019							

Genotype frequency for locus D3S1358, TH01 and CSF1PO for Modern Malay and Orang Asli population in Peninsular Malaysia.

Locus: D3S1358					Locus: TH01					Locus: CSF1PO				
Genotype	MM		OA		Genotype	MM		OA		Genotype	MM		OA	
	P	SD	P	SD		P	SD	P	SD		P	SD	P	SD
14,12			0.018	0.035	6,6	0.029	0.033			9,7	0.010	0.019		
15,12			0.158	0.095	7,6	0.088	0.055	0.018	0.035	10,7	0.010	0.019		
15,13	0.010	0.019			7,7	0.098	0.058	0.018	0.035	10,10	0.010	0.019		
15,14	0.020	0.027	0.158	0.095	8,6	0.010	0.019	0.088	0.074	11,7	0.010	0.019		
15,15	0.059	0.046	**0.211**	0.106	8,7	0.059	0.046	0.070	0.066	11,10	0.196	0.077	0.053	0.058
16,12			0.070	0.066	8,8	0.020	0.027	0.105	0.080	11,11	0.147	0.069	0.053	0.058
16,14	0.039	0.038	0.070	0.066	9,6	0.118	0.063	0.035	0.048	12,9	0.010	0.019		
16,15	0.157	0.071	0.158	0.095	9,7	0.128	0.065	**0.123**	0.085	12,10			0.105	0.080
16,16	**0.167**	0.072	0.018	0.035	9,8	0.069	0.049	0.053	0.058	12,11	0.137	0.068	**0.298**	0.119
17,12			0.018	0.035	9,9	**0.147**	0.069	0.123	0.085	12,12	**0.206**	0.078	0.281	0.117
17,14	0.049	0.042	0.018	0.035	9.3,6	0.020	0.027	0.035	0.048	13,10	0.108	0.060		
17,15	**0.167**	0.072	0.053	0.058	9.3,7	0.069	0.049	0.053	0.058	13,11	0.039	0.038	0.035	0.048
17,16	0.137	0.067	0.018	0.035	9.3,8	0.010	0.019	0.018	0.035	13,12	0.029	0.033	0.140	0.090
17,17	0.039	0.038			9.3,9	0.059	0.046	0.053	0.058	14,11			0.018	0.035
18,14	0.010	0.019			9.3,9.3	0.010	0.019	0.018	0.035	14,12	0.010	0.019	0.018	0.035
18,15	0.029	0.033			10,6	0.020	0.027							
18,16	0.059	0.046	0.035	0.048	10,7	0.020	0.027	0.053	0.058					
18,17	0.0294	0.033			10,8	0.010	0.019	0.053	0.058					
18,18	0.0098	0.019			10,9	0.010	0.019	0.018	0.035					
19,15	0.0098	0.019			10,9.3	0.010	0.033	0.035	0.048					
19,18	0.0098	0.019			10,10		0.055	0.035	0.048					

Genotype frequency for locus D5S818, D8S1179 and D13S317 for Modern Malay and Orang Asli population in Peninsular Malaysia.

Locus: D5S818					Locus: D8S1179					Locus: D13S317				
Genotype	MM		OA		Genotype	MM		OA		Genotype	MM		OA	
	P	SD	P	SD		P	SD	P	SD		P	SD	P	SD
9,7	0.010	0.019			10,10	0.020	0.027			8,7			0.018	0.035
9,9			0.088	0.074	11,10	0.039	0.038			8,8	0.059	0.046	**0.193**	0.102
10,7	0.010	0.019			12,10	0.010	0.019	0.018	0.035	9,8	0.049	0.042	0.035	0.048
10,9	0.020	0.027	0.088	0.074	12,11	0.010	0.019			9,9	0.029	0.033	0.018	0.035
10,10	0.069	0.049	0.053	0.058	12,12	0.020	0.027	0.036	0.049	10,8	0.078	0.052	**0.193**	0.102
11,7	0.010	0.019			13,10	0.039	0.038	0.055	0.060	10,9	0.059	0.046		
11,9	0.039	0.038	0.053	0.058	13,11	0.039	0.038			10,10	0.020	0.027		
11,10	0.167	0.072	**0.105**	0.080	13,12	0.059	0.046	0.091	0.076	11,8	**0.157**	0.071	0.175	0.099
11,11	0.049	0.042	0.053	0.058	13,13	0.020	0.027	0.036	0.049	11,9	0.069	0.049	0.035	0.048
12,7	0.010	0.019			14,10	0.029	0.033			11,10	0.039	0.038		
12,9	0.029	0.033	0.035	0.048	14,11	0.049	0.042			11,11	0.069	0.049		
12,10	**0.197**	0.077	0.070	0.066	14,12	0.069	0.049	0.036	0.049	12,7	0.010	0.019		
12,11	0.157	0.071	0.035	0.048	14,13	0.059	0.046	0.109	0.082	12,8	0.078	0.052		
12,12	0.020	0.027			14,14	0.049	0.042	0.055	0.060	12,9	0.029	0.033		
13,9			0.035	0.048	15,10	0.049	0.042			12,10	0.039	0.038		
13,10	0.088	0.055	0.018	0.035	15,11	0.010	0.019	0.036	0.049	12,11	0.088	0.055		
13,11	0.078	0.052	0.035	0.048	15,12	0.039	0.038	**0.127**	0.088	12,12	0.049	0.042		
13,12	0.049	0.042	0.053	0.058	15,13	0.059	0.046	**0.127**	0.088	13,8	0.020	0.027	0.035	0.048
13,13	0.010	0.019			15,14	**0.108**	0.060	0.055	0.060	13,9	0.010	0.019	0.018	0.035
14,10	0.010	0.019	0.035	0.048	15,15	0.010	0.019	0.073	0.069	13,10	0.010	0.019	0.018	0.035
14,11			0.018	0.035	16,10	0.020	0.027			13,11	0.010	0.019	0.018	0.035
14,13			0.018	0.035	16,11	0.020	0.027			13,12	0.010	0.019	0.018	0.035
15,10	0.010	0.019	0.053	0.058	16,12	0.020	0.027			13,13			0.035	0.048
15,11			0.070	0.066	16,13	0.049	0.042	0.018	0.035	14,7			0.018	0.035
15,12			0.018	0.035	16,14	0.049	0.042			14,8			0.053	0.058
15,13			0.035	0.048	16,15	0.029	0.033	0.018	0.035	14,10			0.035	0.048
15,15			0.035	0.048	17,12			0.018	0.035	14,11	0.020	0.027	0.035	0.048
					17,13	0.010	0.019	0.073	0.069	14,13			0.035	0.048
					17,14	0.010	0.019	0.018	0.035					
					17,15	0.010	0.019							

Allele frequency and forensic parameters for locus D5S818, D8S1179 and D13S317 for Modern Malay and Orang Asli population in Peninsular Malaysia.

Locus: D5S818					Locus: D8S1179					Locus: D13S317				
Allele	MM		OA		Allele	MM		OA		Allele	MM		OA	
	P	SD	P	SD		P	SD	P	SD		P	SD	P	SD
7	0.020	0.019			10	0.113	0.043	0.035	0.034	7	0.005	0.010	0.018	0.024
9	0.050	0.030	0.193	0.072	11	0.083	0.038	0.018	0.025	8	0.250	0.059	**0.465**	0.092
10	**0.304**	0.063	**0.237**	0.078	12	0.123	0.045	0.175	0.071	9	0.137	0.047	0.061	0.044
11	0.275	0.061	0.211	0.075	13	0.176	0.052	**0.263**	0.082	10	0.132	0.046	0.123	0.060
12	0.225	0.057	0.105	0.056	14	**0.235**	0.058	0.158	0.068	11	**0.260**	0.060	0.132	0.062
13	0.118	0.044	0.096	0.054	15	0.162	0.051	0.246	0.080	12	0.176	0.052	0.018	0.024
14	0.005	0.010	0.035	0.034	16	0.093	0.040	0.018	0.025	13	0.029	0.023	0.096	0.054
15	0.005	0.010	0.123	0.060	17	0.015	0.017	0.053	0.042	14	0.010	0.014	0.088	0.052
Forensic Parameters														
PM	0.107		0.060		PM	0.050		0.079		PM	0.072		0.120	
PD	0.893		0.940		PD	0.950		0.921		PD	0.928		0.880	
PIC	0.73		0.80		PIC	0.82		0.77		PIC	0.77		0.70	
PE	0.701		0.548		PE	0.760		0.599		PE	0.553		0.517	
TPI	3.40		2.19		TPI	4.25		2.50		TPI	2.22		2.04	
Homo	0.147		0.228		Homo	0.118		0.20		Homo	0.225		0.246	
Hetero	0.853		0.772		Hetero	0.882		0.80		Hetero	0.775		0.754	
p-values	0.752		0.493		*p-values*	0.938		0.660		*p-values*	0.878		0.129	

PM=Probability of match
PD=Power of Discrimination
PIC=Polymorphism Information Content
TPI=Typical Paternity Index
PE=Power of Exclusion
Homo=Homozygosity
Hetero=Heterozygosity

Allele frequency and forensic parameters for locus D7S820, D16S539 and TPOX for Modern Malay and Orang Asli population in Peninsular Malaysia.

Locus: D7S820					Locus: D16S539					Locus: TPOX				
Allele	**MM**		**OA**		**Allele**	**MM**		**OA**		**Allele**	**MM**		**OA**	
	P	**SD**	**P**	**SD**		**P**	**SD**	**P**	**SD**		**P**	**SD**	**P**	**SD**
7	0.020	0.019			8	0.010	0.014	0.053	0.041	8	**0.574**	0.068	0.421	0.091
8	0.186	0.054	0.193	0.074	9	0.221	0.057	0.149	0.065	9	0.142	0.048	0.079	0.050
9	0.039	0.027			10	0.137	0.047	0.219	0.076	10	0.029	0.023	0.053	0.041
10	0.225	0.058	0.149	0.067	11	**0.284**	0.062	**0.298**	0.084	11	0.235	0.058	**0.447**	0.091
11	**0.324**	0.065	**0.421**	0.093	12	0.196	0.054	0.184	0.071	12	0.020	0.019		
12	0.176	0.053	0.184	0.073	13	0.123	0.045	0.096	0.054					
13	0.025	0.022	0.018	0.025	14	0.025	0.021							
Forensic Parameters														
PM	0.099		0.150		PM	0.074		0.084		PM	0.219		0.211	
PD	0.901		0.850		PD	0.926		0.916		PD	0.781		0.789	
PIC	0.74		0.65		PIC	0.77		0.76		PIC	0.54		0.54	
PE	0.678		0.379		PE	0.518		0.645		PE	0.277		0.195	
TPI	3.16		1.50		TPI	2.04		2.85		TPI	1.21		1.02	
Homo	0.158		0.333		Homo	0.245		0.175		Homo	0.412		0.491	
Hetero	0.842		0.667		Hetero	0.755		0.825		Hetero	0.588		0.509	
p-values	0.593		0.087		*p-values*	0.529		0.735		*p-values*	0.649		0.331	

PM=Probability of match
PD=Power of Discrimination
PIC=Polymorphism Information Content
TPI=Typical Paternity Index
PE=Power of Exclusion
Homo=Homozygosity
Hetero=Heterozygosity

Allele frequency and forensic parameters for locus D3S1358, TH01 and CSF1PO for Modern Malay and Orang Asli population in Peninsular Malaysia.

Locus: D3S1358					Locus: TH01					Locus: CSF1PO				
Allele	**MM**		**OA**		**Allele**	**MM**		**OA**		**Allele**	**MM**		**OA**	
	P	**SD**	**P**	**SD**		**P**	**SD**	**P**	**SD**		**P**	**SD**	**P**	**SD**
12			0.132	0.062	6	0.157	0.050	0.088	0.052	7	0.015	0.017		
13	0.005	0.010			7	0.279	0.062	0.175	0.070	8			0.053	0.041
14	0.059	0.032	0.132	0.062	8	0.098	0.041	0.246	0.079	9	0.010	0.014	0.149	0.065
15	0.255	0.060	**0.474**	0.092	9	**0.338**	0.065	**0.263**	0.081	10	0.201	0.055	0.219	0.076
16	**0.363**	0.066	0.193	0.072	9.3	0.093	0.040	0.114	0.058	11	**0.392**	0.067	**0.298**	0.084
17	0.230	0.058	0.053	0.041	10	0.034	0.025	0.114	0.058	12	0.304	0.063	0.184	0.071
18	0.078	0.037	0.018	0.024						13	0.074	0.036	0.096	0.054
19	0.010	0.014								14	0.005	0.010		
Forensic Parameters														
PM	0.114		0.135		PM	0.088		0.074		PM	0.142		0.206	
PD	0.886		0.865		PD	0.912		0.926		PD	0.858		0.794	
PIC	0.70		0.66		PIC	0.73		0.78		PIC	0.66		0.55	
PE	0.469		0.548		PE	0.422		0.431		PE	0.485		0.379	
TPI	1.82		2.19		TPI	1.65		1.68		TPI	1.89		1.50	
Homo	0.275		0.228		Homo	0.304		0.298		Homo	0.265		0.333	
Hetero	0.725		0.772		Hetero	0.696		0.702		Hetero	0.735		0.667	
p-values	0.499		0.713		*p-values*	0.627		0.113		*p-values*	0.191		0.948	

PM=Probability of match
PD=Power of Discrimination
PIC=Polymorphism Information Content
TPI=Typical Paternity Index
PE=Power of Exclusion
Homo=Homozygosity
Hetero=Heterozygosity

Allele frequency and forensic parameters for locus D18S51, PENTA D and vWA for Modern Malay and Orang Asli population in Peninsular Malaysia.

Locus: D18S51					Locus: Penta D					Locus: vWA				
Allele	MM		OA		Allele	MM		OA		Allele	MM		OA	
	P	SD	P	SD		P	SD	P	SD		P	SD	P	SD
10	0.005	0.010	0.009	0.017	7	0.020	0.019			14	0.211	0.056	0.184	0.071
10.2			0.009	0.017	8	0.029	0.023	0.009	0.017	15	0.034	0.025		
11	0.005	0.010			9	**0.382**	0.067	0.272	0.082	16	0.142	0.048	0.140	0.064
12	0.103	0.042	0.070	0.047	10	0.181	0.053	0.149	0.065	17	**0.245**	0.059	**0.447**	0.091
13	0.098	0.041	0.035	0.034	11	0.132	0.046	**0.360**	0.088	18	0.240	0.059	0.167	0.068
14	0.157	0.050	0.114	0.058	12	0.142	0.048	0.044	0.038	19	0.088	0.039	0.061	0.044
14.2			0.009	0.017	13	0.078	0.037	0.061	0.044	20	0.039	0.027		
15	**0.250**	0.059	**0.474**	0.092	14	0.029	0.023	0.061	0.044					
16	0.157	0.050	0.237	0.078	15			0.044	0.038					
17	0.093	0.040												
18	0.054	0.031	0.009	0.017										
19	0.034	0.025	0.009	0.017										
20	0.020	0.019												
21	0.010	0.014	0.026	0.029										
23	0.010	0.014												
24	0.005	0.010												
Forensic Parameters														
PM	0.046		0.121		PM	0.083		0.104		PM	0.074		0.120	
PD	0.954		0.879		PD	0.917		0.896		PD	0.926		0.880	
PIC	0.84		0.66		PIC	0.74		0.73		PIC	0.78		0.68	
PE	0.779		0.431		PE	0.553		0.548		PE	0.643		0.404	
TPI	4.64		1.68		TPI	2.22		2.19		TPI	2.83		1.58	
Homo	0.108		0.298		Homo	0.225		0.228		Homo	0.176		0.316	
Hetero	0.892		0.702		Hetero	0.775		0.772		Hetero	0.824		0.684	
p-values	0.578		0.524		*p-values*	0.721		0.422		*p-values*	0.585		0.816	

Allele frequency and forensic parameters for locus D21S11, FGA and PENTA E for Modern Malay and Orang Asli population in Peninsular Malaysia.

Locus: D21S11					Locus: FGA					Locus: PENTA E				
Allele	MM		OA		Allele	MM		OA		Allele	MM		OA	
	P	SD	P	SD		P	SD	P	SD		P	SD	P	SD
28	0.059	0.033	0.053	0.042	18	0.005	0.010			5	0.039	0.027		
29	**0.240**	0.060	**0.351**	0.089	19	0.074	0.036	0.105	0.056	7	0.015	0.017		
30	0.225	0.058	0.105	0.057	20	0.059	0.032	0.158	0.067	9	0.029	0.023		
30.2	0.039	0.027	0.114	0.059	20.2	0.010	0.014			10	0.005	0.010	0.026	0.029
31	0.142	0.049	0.009	0.018	21	0.147	0.049	**0.203**	0.074	11	**0.255**	0.060	**0.412**	0.090
31.2	0.083	0.039	0.175	0.071	21.2	0.020	0.019	0.053	0.041	12	0.176	0.052		
32	0.034	0.025	0.009	0.018	22	**0.260**	0.060	0.158	0.067	13	0.083	0.038	0.053	0.041
32.2	0.113	0.044	0.096	0.055	22.2	0.029	0.023	0.096	0.054	14	0.074	0.036	0.228	0.077
33	0.005	0.010			23	0.147	0.049	0.096	0.054	15	0.059	0.032	0.044	0.038
33.2	0.039	0.027	0.079	0.050	24	0.108	0.043	0.061	0.044	16	0.059	0.032		
34.2	0.005	0.010			24.2	0.005	0.010	0.009	0.017	17	0.044	0.028	0.053	0.041
					25	0.074	0.036	0.053	0.041	18	0.059	0.032	0.070	0.047
					25.2	0.005	0.010			19	0.034	0.025	0.044	0.038
					26	0.039	0.027			20	0.034	0.025	0.044	0.038
					26.2	0.005	0.010			21	0.015	0.017		
					27	0.005	0.010			22	0.010	0.014		
					28	0.005	0.010			23			0.026	0.029
										24	0.005	0.010		
Forensic Parameters														
PM	0.067		0.074		PM	0.040		0.050		PM	0.034		0.091	
PD	0.933		0.926		PD	0.960		0.950		PD	0.966		0.909	
PIC	0.82		0.78		PIC	0.84		0.85		PIC	0.86		0.73	
PE	0.709		0.599		PE	0.757		0.785		PE	0.760		0.431	
TPI	3.50		2.50		TPI	4.21		4.75		TPI	4.25		1.68	
Homo	0.143		0.20		Homo	0.119		0.105		Homo	0.118		0.298	
Hetero	0.857		0.80		Hetero	0.881		0.895		Hetero	0.882		0.702	
p-values	0.010		0.522		*p-values*	0.651		0.201		*p-values*	0.730		0.129	

Phylogenetic tree of mtDNA haplogroups found within the Modern Malay and Orang Asli populations in Peninsular Malaysia. The tree was rooted using the L3 haplogroup. Classifications were based on the coding and non-coding region polymorphisms, as shown under each branch. The recurrent mutations are underlined.