UNIVERSITI SAINS MALAYSIA

PUSAT PENGAJIAN SAINS KESIHATAN

(SCHOOL OF HEALTH SCIENCES, UNIVERSITI SAINS MALAYSIA)

Classification of Haplogroup C and Haplogroup D in Modern Malay and Chinese population of Peninsular Malaysia

Dissertation submitted in partial fulfillment of the requirements for the

Degree of Bachelor of Health Sciences (Forensics)

EDINUR HISHAM BIN ATAN

School of Health Sciences

Universiti Sains Malaysia

Health Campus

16150 Kubang Kerian, Kelantan

Malaysia

MEI 2006

CERTIFICATE

This certify that the dissertation entitled

"Classification of Haplogroup C and Haplogroup D in Modern Malay and Chinese population of Peninsular Malaysia"

is a bonafide record of research work done by

Mr. Edinur hisham bin Atan

During the period from December 2005 to April 2006

Under our supervision

Signature of supervisor:.....

Name and address of supervisor: Dr. Zafarina Zainuddin
Lecturer
School of Health Sciences
Universiti Sains Malaysia
Health Campus
16150 Kubang Kerian, Kelantan

Date: 9 MEI 2006

ACKNOWLEDGEMENT

First of all, I would like to thank god because at last I had finish doing this thesis. Although there are many problems I face on the process of making this paper, I face it bravely in order to make it become reality.

I am indebted to my supervisor, Dr. Zafarina Zainuddin because she teaches me a lot from the selection of this topic to the process of making this paper. Her ideas make this paper become more interesting and systematic. I also would like to give acknowledgement to my parent because they teach me how to be a good student and son. I'm really appreciate them because they always beside me whether I'm happy or sad.

Here also, I would like to acknowledge to all my lecturers in Universiti Sains Malaysia (USM) Kubang Kerian because they create me to this level and I'm really proud to have them as my teachers.

In addition, I also would like to acknowledge to my entire friend either in science university of Malaysia or in other university. Thanks for your support and always cheer me up. Thanks my friend.

My efforts to write this seminar paper would be failing without support and guidance of all above people. Thanks and I hope we can be together for ever.

LIST OF CONTENT **PAGE**

LIST OF FIGURE **PAGE**

LIST OF TABLE **PAGE**

1. ABSTRACT

Mitochondrial DNA analysis is carried out for 45 samples from modern Malay and 44 samples from modern Chinese populations of Peninsula Malaysia. The mitochondrial DNA coding region variation of these samples was examined using restriction fragment length polymorphism (RFLP) method. From the analysis of *Alu*1 site gain nucleotide position 10397, 37.78% of modern Malay population of peninsular Malaysia belongs to superhaplogroup M while 47.73% of the modern Chinese belong to superhaplogroup M.

Superhaplogroup M is further divided into several haplogroup, which are defined by the presence of specific RFLP markers. Haplogroup C and Haplogroup D are determined by the presence of Alu1 restriction site at the nucleotide position 13 262 and the absence of *Alu*1 restriction site at the nucleotide position 5 176. For Modern Malay population, only 8.89% of the sample belongs to the haplogroup D while none belong to the haplogroup C. For modern Chinese population of peninsular Malaysia, 4.65% of the samples belong to haplogroup C and 6.82% of the samples belong to the haplogroup D.

2. INTRODUCTION

2.1 Human mtDNA

Human mtDNA is 16,569 bp in length located in the mitochondrial organelle (Chin-Yuan, 2001). Mitochondria are the main sites of biological energy generation in eukaryotes (Mark *et al.*, 2004). Mitochondria contain two membranes, separated by a space. Inside the space enclosed by the inner membrane is the matrix. This appears moderately dense and one may find strands of DNA, ribosome, or small granules in the matrix. The mtDNA region is 15,383 bp in length and encodes 22 tRNAs, 2 rRNAs, and 13 structural proteins for oxidative phosphorylation (Chin-Yuan, 2001).

Mitochondria are the power stations of the cell, providing energy for every specialized cell function in the form of ATP or heat. Essential parts of the oxidative phosphorylation apparatus to generate energy are encoded by the mitochondrial genome (Rebecca L. Cann, 1987). The food that we eat is oxidized to produce high-energy electrons that are converting to form energy in the form of high energy phosphate bone called triphosphate or ATP (Gwen V. Child, 1998). Pyruvate oxidation, citric acid cycle, electron transport and oxidative phosphorylation also take place in mitochondria (Mark *et al.*, 2004).

Mitochondrial DNA is passed from one generation to the next, essentially unchanged, solely through the maternal line of a family (Thomas, 1997). Thus, the mtDNA sequences obtained from maternally related individuals, such as a brother and a sister or a mother and a daughter, will exactly match each other in the absence of a mutation (Alice, 1999)

The mitochondrial genome is exclusively maternally inherited and demonstrates no recombination events, but does exhibit high mutation and mutation fixation rates (Sabine Hofmann *et al.*, 1997). It exhibits a few interesting features which make it a suitable tool for phylogenetic and pathogenetic studies (Sabine Hofmann, 1997). Mitochondrial DNA gives a magnified view of the diversity present in the human gene pool because mutations accumulate in this DNA several times faster than in the nuclear DNA. The rate of sequence evolution in mtDNA is 10-20 times higher than in nuclear genome and consequently any two mtDNA may differ by 10-66 nucleotides from each other (Zeviani *et al.*, 1998).

Mitochondrial DNA is maternally inherited maternally and does not undergo recombination. It is therefore useful as a tool for identification of maternally related individual. Mitochondrial DNA also found in high copy number in each cells. There are about 1016 mtDNA molecules within a typical human cell and they are usually identical to one another 17-19 (Rebecca, 1987). Complete mitochondrial genome sequence has gained importance in resolving phylogenies and understanding human evolution (Rajkumar *et al.*, 2005). By convention, human mtDNA sequences are described using the first complete published mtDNA sequence as a reference (Anderson *et al.*, 1981). The revised version where a few corrections have been made is known as Cambridge Reference Sequence (CRS) (Andrews *et al.*, 1999).

2.1.1 Organization of mtDNA

Mitochondria occur in various forms depending on the cell or tissue type and even within single cells, reflecting the variety of cellular functions localized to these organelles (Yaffe 1999, Collins *et al.*, 2002). The mitochondrial membranes can be distinguished from other cellular membranes due to the characteristic double-membrane structure (Frey & Mannella, 2000), Perkins & Frey 2000, Frey *et al.*, 2002). According to the current model, the inner-membrane cristae structures are typically attached to the mitochondrial outer membrane by very thin connections only (Sjöstrand, 1953).

The human mtDNA genome is approximately 16,569 bases in length and has two general regions: the coding region and the control region. The coding region is responsible for the production of various biological molecules involved in the process of energy production in the cell. The control region is responsible for regulation of the mtDNA molecule. Two regions of mtDNA within the control region have been found to be highly polymorphic, or variable, within the human population (Greenberg *et al.*, 1983). These two regions are known as hypervariable region I (HVR-1), which has an approximate length of 342 base pairs (bp), and Hypervariable Region II (HVR-2), which has an approximate length of 268 bp (R. Isenberg, 1999).

The control region is the primary non-coding region, and is responsible for the regulation of heavy (H) and light (L) strand transcription and of H-strand replication. The analysis of the hypervariable control region of human mitochondrial DNA (mtDNA) was proven to be a useful tool for forensic scientists, population and evolutionary biologists and anthropologists (Miller *et al.*, 2000).

2.1.2 Advantages of mtDNA in Forensic DNA analysis

A special feature of mtDNA is the high copy number in each cell. The high copy number of mtDNA gives the forensic scientist a better chance obtaining DNA profile from limited or severely degraded sample (R. Isenberg, 2002).

The location and structure of mtDNA protect it from degradation when exposed to the environment. Mitochondrial DNA is buried deep within the cell and has a circular structure, which protects it from deterioration (Isenberg, 2002). This feature is very important for the biological sample which exposed to adverse environmental and prevent it from degradation.

As a molecular marker, mitochondrial DNA has many advantages. It evolves faster than nuclear DNA (Brown *et al.* 1982), probably due to inefficient replication repair (Clayton 1984). Different regions of the mitochondrial genome evolve at different rates and allowing suitable regions to be chosen for the question under study (Saccone *et al.*, 1991). Mitochondrial DNA does not recombine (Hayashi *et al.*, 1985), though some evidence of recombination events has recently been reported (Eyre-Walker *et al.*, 1999, Hagelberg *et al.*, 1999). In this regard, three hypervariable regions (HVRs) in the D-loop have been identified and PCR-generated DNA sequencing data have been used for forensic purposes (Tzen *et al.*, 2001).

Mitochondrial DNA is maternally inherited in most species (exceptions with paternal leakage including mice, Gyllesten *et al.*, 1991; biparental inheritance in marine mussels,

Zouros *et al.,* 1992). It can be used for the identification of skeletal remains if mtDNA of the same maternal lineage can be obtained (Tzen, 2001).

2.2 Mitochondrial DNA Analysis

Mammalian cells have two types of genome, nuclear genome and mitochondrial genome. Both genomes are independent of each other. Mitochondrial genome is semi-autonomously functioning organelle containing a resident genome that undergoes replication, translation and transcription of their own DNA. Each organelle in cells posses at least 2-10 copies of mitochondrial DNA and therefore each cell may contains more than 1000 copies of mitochondrial DNA. This feature is very important in forensic DNA analysis which usually confront with degradable sample.

Population studies use restriction amplified fragment length polymorphism (RFLP) technique to analyze mitochondrial DNA coding region. RFLP variation also revealed continent-specific polymorphism for classifying mtDNA (Michael *et al.*, 2005). Mitochondrial DNA coding region has a number of stable polymorphic sites compare to control region which undergo a high mutation rate. Amplification of mitochondrial DNA coding region fragment followed by restriction digestion is used to classify mtDNA haplogroup.

2.3 Application of mtDNA In Population Studies

Variability of human mitochondrial DNA has provided valuable data about genetic past of human maternal lineage. Analysis of the frequency, variation and distribution of mitochondrial DNA haplotype are used to evaluate current model concerning the process of colonization of the world (Piia Serk, 2004). Uniparental inheritance and high mutation rate have led to mtDNA lineages (haplogroups), which are defined by ancient polymorphisms and characterized by considerable variation (Anna *et al.*, 2004). MtDNA studies hit the popular consciousness in 1987, as a result of the publicity surrounding the debate on modern human origins and the African "mitochondrial Eve" (Cann *et al.*, 1987).

The mtDNA mutations have accumulated sequentially along radiating maternal lineages and now characterize human populations in different geographical regions of the world. Classification of the genotypes has been made based on differences found in restriction fragment analysis of the coding region or in the sequence of the hypervariable segment I. Both methods have shortcomings, as the former may not detect all the important polymorphisms and the latter makes use of a segment containing hypervariable nucleotide positions (Finnila *et al.*, 2000). Restriction fragment length polymorphism (RFLP) studies of mtDNA coding regions have been used to classify such lineages (Torroni *et al.*, 1996). The European population is almost exclusively distributed among the nine haplogroups designated as H, I, J, K, T, U, V,W and X, whereas haplogroups A,B, C, D, E, F, G and certain subclusters of superhaplogroups M and N are characteristic to Asian populations; haplogroups A, B, C and D are more familiar to

native Americans while haplogroup L1, L2 and L3 are more specific to African populations (Niemi and Moilanen, 2004)

Restriction fragment length polymorphism (RFLP) studies of mtDNA coding regions have been used to classify such lineages or haplogroup (Torroni *et al.,* 1996). The mutation that have struck through human history trace the maternal genealogy and through this genealogy we can attempt to deduce something about prehistoric process (Macaulay *et al.,* 1999) The structure of the tree was rather star like, there was a single, central haplotype, shared among individuals from all over the world, that radiated other types, some of which were population specific. The central (so-called "universal") haplotype was assumed to be the type of the most recent common ancestor of all extant mtDNAs in the world. This suggested that all human populations might have shared a common evolutionary history for a very long time (Richards and Macaulay, 2000)

Table 2.1: Common specific mtDNA haplogroup all around the globe.

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*HI/*Mbo*I
	J	13704 –*BstN*I; 16065 –*Hinf*I
	K	9052 –*Hae*II/*Hha*I
Asia	A	663 +*Hae*III
	B	8271-8291 9bp deletion; 16517 +*Hae*III
	F	12406 –*Hpa*I/*Hinc*II; 16517 +*Hae*III
	M	10394 +*Dde*I; 10397 +*Alu*I
	M-C	13259 –/13262 +*Hinc*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 +*Hinc*II/*Alu*I
	M-D (Amerinds)	5176 –*Alu*I

Source: www.mitomap.org

3. Literature review:

Human mtDNA is a non-recombining molecule with maternal inheritance and practically haploid genetics. Differences between mtDNA sequences are only due to mutation (Nicole *et al.*, 2001). Two aspects of the mitochondrial DNA make it particularly useful in human evolutionary studies. Firstly, contrast to the nuclear encoded gene, which shows Mandelian inheritance mtDNA has been shown to be maternally inherited (Giles *et al.*, 1980) and secondly, the mtDNA sequence evolution rate is much higher than that of average nuclear gene (Miyata *et al.*, 1982; Wallace *et al.*, 1987). Coding region sequences evolved according to a molecular clock, but that control region evolution was not clock-like (Torroni *et al.*, 2001). The rate of sequence evolution in mtDNA is 10–20 times higher than that in the nuclear genome, and, consequently, any two mtDNAs may differ by 10–66 nt from each other (Zeviani *et al.*, 1998; Chinnery et *al.*, 1999).

Mitochondrial DNA (mtDNA) can be used to trace maternal ancestry. The geographic distribution and variation of mtDNAs can be highly informative in defining potential range expansions and migration routes in the distant past (Lluis *et al.*,1990). Complete mitochondrial genome sequencing has gained importance in resolving phylogenies and understanding human evolution (Revathi *et al.*, 2005). Human mtDNA has fully sequenced for more than 20 years ago and make it possible to study variation in human population (Anderson *et al.*, 1981).

In recent years, different methods based on gel electrophoresis of PCR-generated products have been developed for detecting single-base mismatches in DNA, including

single-strand conformation polymorphism analysis (Thomas *et al.,* 1994), denaturing gradient gel electrophoresis (Gross *et al.,* 1994), low-stringency single-specific primer PCR (Pena et al., 1994), and mutation detection enhancement gel matrix (Alonso *et al.,* 1996). Conformation-sensitive gel electrophoresis (CSGE) is based on the separation of heteroduplexes containing single-base-pair mismatches from homoduplexes in a polyacrylamide gel (Ganguly *et al.,* 1993).

Restriction fragment length polymorphism (RFLP) studies of mtDNAs from a wide range of human populations have revealed a number of stable polymorphic sites in the mtDNA coding regions. These define related groups of mtDNAs called haplogroups. Most of the mutations observed in both mtDNA coding and control regions in modern human populations have occurred on these preexisting haplogroups and define the individual mtDNA types or haplotypes (Torroni *et al.,* 1993a; Graven *et al.,* 1995). RFLP have revealed a number of stable polymorphic sites that define mtDNA haplogroups, most of which have been shown to be continent specific (Wallace, 1994). RFLP technique used to study human population only allow 20% of mitochondria DNA sequence to be examined. Therefore, sequencing of hypervariable site in D-loop region help in classification of mitochondrial DNA into haplogroup. A good correlation has been obtained between the RFLP data and HVS-I sequence data (Torroni *et al.,* 1996, 1998; Macaulay *et al.,* 1999).

The majority of haplogroups have been shown to be continent-specific. For example, in Africa, haplogroup L encompasses between 70 and 100% of the sub-Saharan mtDNAs (Chen *et al.,* 1995; Grave *et al.* 1995). In Asia, 55% of East Asiana and Siberian mtDNAs

are members of superhaplogroup M (Ballinger *et al.,* 1992; Torroni *et al.,* 1993b,c; Chen *et al.* 1995; Wallace 1995). Haplogroup M is subdivided into smaller subhaplogroups designated C, D, G and E. Most of the remaining Asian mtDNAs are encompassed by haplogroups A, B and F (Torroni *et al.,* 1994). The presence (20%) of the 'Asian' mtDNA haplogroup M, defined by the 10400 C to T transition, is unique in eastern Africa (Lluis *et al.,* 1999)

Among Native Americans, only four Asian haplogroups (A, B, C and D) are observed, thus indicating that these haplogroups predated the colonization of the Americas (Schurr *et al.,* 1990; Ward *et al.,* 1991; Horai *et al.,* 1993; Torroni e*t al.* 1993a, 1994a,d). Ten haplogroups (H, I, J, K, M, T, U, V, W, and X) encompass almost all mtDNAs from European populations (Torroni *et al.,* 1996). Finally, haplogroups A, B, C, D, E, F, G and M embrace the majority of the lineages described for Asia, Oceania and Native Americans (Nicole *et al.,* 2001). Complete sequencing of 62 mitochondrial DNAs (mtDNAs) belonging (or very closely related) to haplogroup H revealed that this mtDNA haplogroup—by far the most common in Europe—is subdivided into numerous subhaplogroups, with at least 15 of them (H1–H15) identifiable by characteristic mutations (Achilli *et al.,*2004)

The now-emerging mitochondrial DNA (mtDNA) population genomics provides information for reconstructing a well-resolved mtDNA phylogeny and for discerning the phylogenetic status of the subcontinentally specific haplogroups (Kong *et al.,* 2003). Analysis of mitochondrial has provided evidence that our species arose in Africa about 150000 years before present (YBP), migrated out of Africa into Asia about 60,000 to

70000 YBP and into Europe about 40,000 to 50,000 YBP, and migrated from Asia and possibly Europe to Americas about 20000 to 30000 YBP (Wallace *et al.*, 1999). Both archaeology and genetics suggest that modern humans originated 100,000 to 200,000 years ago in Africa (Cann *et al.*, 1987; Stringer, 1990). There is no substantial evidence supporting further spatial dispersal of modern humans earlier than around 50,000 years ago (Kivisild *et al.*, 1999). What happened during this 50,000 years long gap and where did the initial radiation of the Eurasian population take place remains largely an open question.

Two major routes of dispersal have been hypothesized: one through North Africa into the Levant, documented by fossil remains, and one through Ethiopia along South Asia, for which little, if any, evidence exists (Quintana *et al.*, 1999). After coming out of Africa, modern humans first spread to Asia following two main routes. The southern one is represented by haplogroup M and related clades that are overwhelmingly present in India and eastern Asia. The northern one gave a posterior radiation that, through Central Asia, again reached North and East Asia carrying, among others, the prominent lineages A and B. Later expansions can be detected by the presence of subclades of haplogroup U in India and Europe. There were also returns to Africa, most probably from the same two routes. The return from India could be detected by the presence of derivatives of M in Northeast Africa, and the arrival of Caucasoids by the existence of a subclade of haplogroup U that, today, is mainly confined to Northwest Africa (Maca-Meyer *et al.*, 2001)

4. OBJECTIVES

The objectives of this project are:

a) To examine the mtDNA of the modern Malay and Chinese populations in Peninsular Malaysia.

b) To obtain mtDNA profile of modern Malay and Chinese populations in Peninsular Malaysia.

c) To identify and discovered the evolution of modern Malay and Chinese populations in Peninsular Malaysia based on mtDNA profile.