UNIVERSITI SAINS MALAYSIA

UNIVERSITI SAINS MALAYSIA

Hypervariable segment-I (HVS-I) Sequence Analysis Of Chinese Population of Peninsular Malaysia

Dissertation submitted in partial fulfillment for the Degree of Bachelor of Science in Forensic Science

Lee Loong Chuen

School of Health Sciences
Universiti Sains Malaysia
Health Campus
16150, Kubang Kerian, Kelantan
Malaysia

2007

CERTIFICATE

This is to certify that the dissertation entitled

"Hypervariable segment-I (HVS-I) Sequence Analysis

Of Chinese Population of Peninsular Malaysia"

is the bonafide record of research work done by

Ms. Lee Loong Chuen

during the period **December 2006** to **April 2007**

under my supervision.

ACKNOWLEDGEMENTS

I would like to dedicate my appreciation and million of thanks to my family for their continuous moral support. Sincerest thanks to Dr. Zafarina Zainudin, my supervisior for all her valuable help, opinions, information and support, which guided me in completing this research project. Specifically, gratitude is owed to my seniors, Nor Alfarizan Binti Mokhtaruddin and Siti Balkiah Binti Ismail. The DNA samples provided to me are greatly appreciated.

Many thanks to Mrs. Rosniah Yusoff, Ms. Nur Arnida and all laboratory staffs from the forensic laboratory and molecular biology laboratory for their willingness to help me to finished my laboratory work.

My special thanks to all my friends, especially Ayu Suzailiana Binti Muhamad for their support that accompanying me from start tills the end of the project.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

DNA	Deoxyribonucleic acid
mtDNA	Mitochondrial DNA
nDNA	Nuclear DNA
YBP	Years before present
RFLP	Restriction fragment length polymorphisms
PCR	Polymerase chain reaction
tRNA	Transfer ribonucleic acid
rRNA	Ribosomal ribonucleic acid
CR	Control region
HVS-I	Hypervariable segment I
HVS-II	Hypervariable segment II
HVS-III	Hypervariable segment III
bp	Base pair
Kb	Kilobase
np	Nucleoctide position
RE	Restriction endonucleases

ABSTRACT

Throughout Malaysia history, most of the Peninsular Malaysian Chinese are known to be populated by those migrated from the China during the British colonial era (Barbie, 1992). The present study was undertaken to obtain information about the distribution of the existing HVS-I sequence variations in the Chinese population of Peninsular Malaysia. A few studies have used RFLP analysis to characterize the mtDNA of Malaysian Chinese (Ismail, 2006; Mokhtaruddin, 2006; Ballinger *et al.,* 1992). The primary purpose of this study is to provide HVS-I sequence data by sequencing the HVS-I of the 40 Chinese individuals from different parts of Peninsular Malaysia. The secondary purpose is classification of mtDNA haplogroup based on the HVS-I sequence variation by alignment of samples sequence and Cambridge Reference Sequence (CRS) (Andrew *et al.*, 1999).

The samples were predicted to belong to the superhaplogroup M and haplogroup B in relatively high frequency as inferred from the study done by Ballinger *et al.,* (1992). Analysis of HVS-I within the control region (CR) of the mitochondrial genome revealed the presence of eight haplogroups: B, F, M, MC, MD, MG, N, and R. These haplogroups are the characteristics of Southeast Asian populations (Ballinger *et al.,* 1992) and also found in significant proportion in Han Chinese of China (Yao *et al.,* 2002). Result shows that superhaplogroup M (includes MC, MD, and MG) is the most common haplogroup in the studied samples (55%), followed by haplogroup B (18%).

INTRODUCTION

Human mtDNA has become a useful tool in forensic science due to its unique features: maternal inheritance (Giles *et al.*, 1980; Chen *et al.*, 1995a; Herrnstadt *et al.*, 2002; Sutovsky *et al.*, 2004); has no recombination (Lewin, 1999; Kivisild and Villems, 2000; Kumar *et al.*, 2000; Parsons and Irwin, 2000; Elson *et al.*, 2001); shows high rate of mutations of the control region (CR)(Howell *et al.*, 1996) and high copy number in cells (Wallace *et al.*, 1997).

The mtDNA diverges at the rate of 2 – 4% per site per million years (Cann *et al.*, 1987), and evolves 6-17 times faster than nuclear DNA (Brown *et al.*, 1979; Wilson *et al.*, 1985). Rapid evolution of mtDNA has resulted in multiple restriction fragment length polymorphisms (RFLP), and nucleotide variants in CR and coding region (Anderson *et al.*, 1981; Wallace *et al.*, 1987), which are correlated with the ethnic and geographic origin of individuals, presumably because mtDNA mutations have accumulated along with radiation of maternal lineages as women migrated out of Africa into different continents (Cann *et al.*, 1987; Stringer and Andrews 1988; Merriwether *et al.*, 1991; Torroni *et al.*, 1992; Horai *et al.*, 1995).

The classification of mtDNA haplogroups is based on information gained from RFLP analysis of the coding region and from the nucleotide sequences of the hypervariable segment (HVS) in the CR (Ballinger *et al.*, 1992; Torroni *et al.*, 1996; Richards *et al.*, 1998.). Polymorphisms used to determine mtDNA haplogroups occurred 10,000 years ago are today high-prevalence population-specific substitutions and most of them are continent-specific. Haplotypes are sub-clusters of haplogroups, and the polymorphisms that determine them are less prevalent and have occurred more recently (Wallace, 1995).

In this study, HVS-I has been chosen to be sequenced, as it contains more variation per nucleotide than other regions (Lutz *et al.,* 1998; Vigilant *et al.,* 1991), thus would provide more mutation motif for the classification of mtDNA haplogroup. Foster *et al.,* (1996) have estimated the mutation rate in HVS-I as 20,180 years per transition.

The samples of this study composed of Peninsular Malaysian Chinese, which mostly populated by migrants radiating out from China. Migration of Chinese to Southeast Asia was dated 2,000 years ago. The first arrival in Peninsular Malaysia was recorded in 1349. The waves of Chinese migration especially from southern China (included Fujian, Hakkas, Guangdong, Chaozhou, and Hainan) into Malaysia, during the mid-19th to early 20th centuries were swelled in the British colonial era, pulled by the economic opportunities and pushed by the dire conditions in China. The British left Chinese immigration uncontrolled until 1930 (Barbie, 1992). Geographical distribution of Malaysian Chinese in Malaysia is shown by figure 1.

Several studies were carried out to determine the mtDNA haplogroups for Malaysian Chinese (Ismail, 2006; Mokhtaruddin, 2006; Ballinger *et al.,* 1992) using RFLP method, which focused on the variation of coding region of mtDNA. On the other hand, this study is to provide HVS-I sequence data for the Chinese population of Peninsular Malaysia using direct sequencing method.

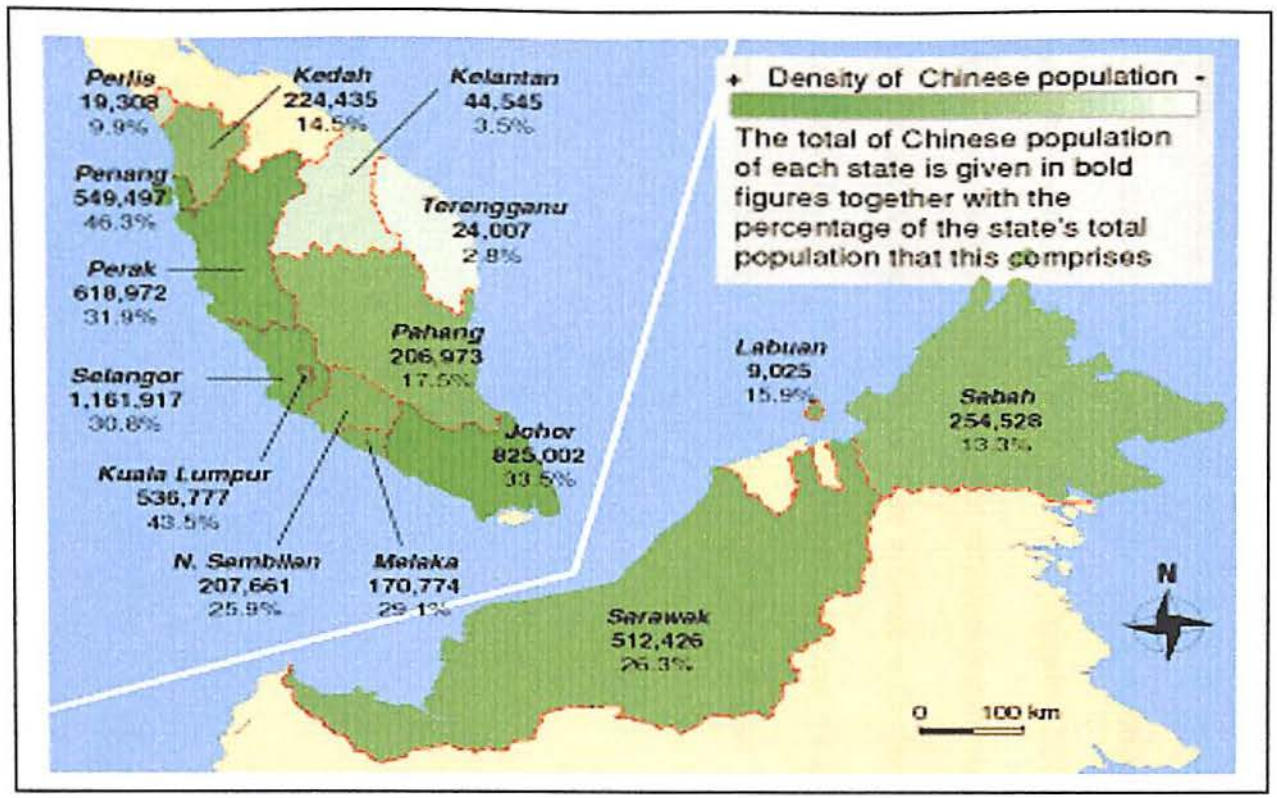

Figure 1: Geographical distribution of the Chinese population in Malaysia.

The density of the Chinese population is indicated by the intensity of green color; being the state comprised the densest Chinese population having the most darken color. Most of the Chinese located at the west peninsular Malaysia, in which Selangor is the only state encompasses more than one million Chinese citizens. In Penang (46.3%), Perak (31.9%), Selangor (30.8%), Kuala Lumpur (43.5%) and Johor (33.5%), more than 30% of the state's total population consists of Chinese population. Only three states be composed of less than 10% of Chinese population in its state, which includes Perlis (9.9%), Kelantan (3.5%) and Terengganu (2.9%).(http://www.encyclopedia.com.my/volume9/dist_chinese.html)

Direct sequencing method is similar to the principle of PCR technique, except for the composition of reaction mixture. The DNA sample is denatured, annealed, and elongated in the same way as PCR, but additional molecules are present in the reactions. These molecules, called dideoxyribonucleotides which behave in the same way as the other nucleotides in the PCR solution, but they contain a fluorescent tag and arrest DNA polymerization when incorporated into a growing DNA strand. As the double-stranded DNA denatured into two single strands, binding of nucleotides occur during elongation phase. Randomly, the strand will bind with a dideoxy base instead of a normal nucleotide and the elongation of the strand will stop. This result in a fragment of double-stranded DNA, which would later subject to capillary electrophoresis that separate the different fragments based on their size.

The samples sequences were then aligned with the Cambridge Reference Sequence (CRS) (Andrew *et al.,* 1999) and the polymorphisms are reported based on the nucleotide variations. Subsequently, each sample is classified into specified haplogroup using the definitive motifs of the haplotypes.

REVIEW OF LITERATURE

ORGANIZATION OF MITOCHONDRIAL DNA

The mtDNA is a circular double-stranded molecule with an extra-nuclear origin. The human mtDNA consists of 16,569 base pairs (bp) and encodes for 22 tRNAs as well as 2 rRNAs. These structural RNAs are used to translate the 13 polypeptides that are subunits of the oxidative phosphorylation (OXPHOS) enzyme complexes, including complex I, III, IV, and V (Lightowlers *et al.*, 1997; Wallace *et al.*, 1999).

The mtDNA encompasses a pyrimidine-rich strand known as the "light strand", and a purine-rich strand known as the "heavy strand". Part of the mtDNA genome that has no known coding function is identified as the control region (CR). The CR contains the 'D-loop' or Displacement loop which describes a structure formed during mtDNA replication which has its origin in this region (Holland and Parsons, 1999). An *Mbo* I restriction site within the major CR has been arbitrarily designated as the origin of the circular DNA, and the base pairs are numbered sequentially proceeding counterclockwise. The CR can be further divided into: HVS-I (positions 16024 - 16365), HVS-II (positions 73-340) and HVS-III (positions 438 - 574) (Greenberg *et al.*, 1983; Wallace *et al.*, 1995; Wallace *et al.*, 1999; Smith, 2002). Forensic mtDNA examinations are performed using these regions because of the high degree of variability found among individuals.

The first mtDNA sequence was reported by Anderson *et al.*, (1981) using placenta of an individual with European ancestry. Re-analysis of original placental material by Andrews *et al.*, (1999) found 11 nucleotides that differed from Anderson *et al.*, (1981) sequence. This Cambridge Reference Sequence (CRS) is now been accepted as the standard for comparison.

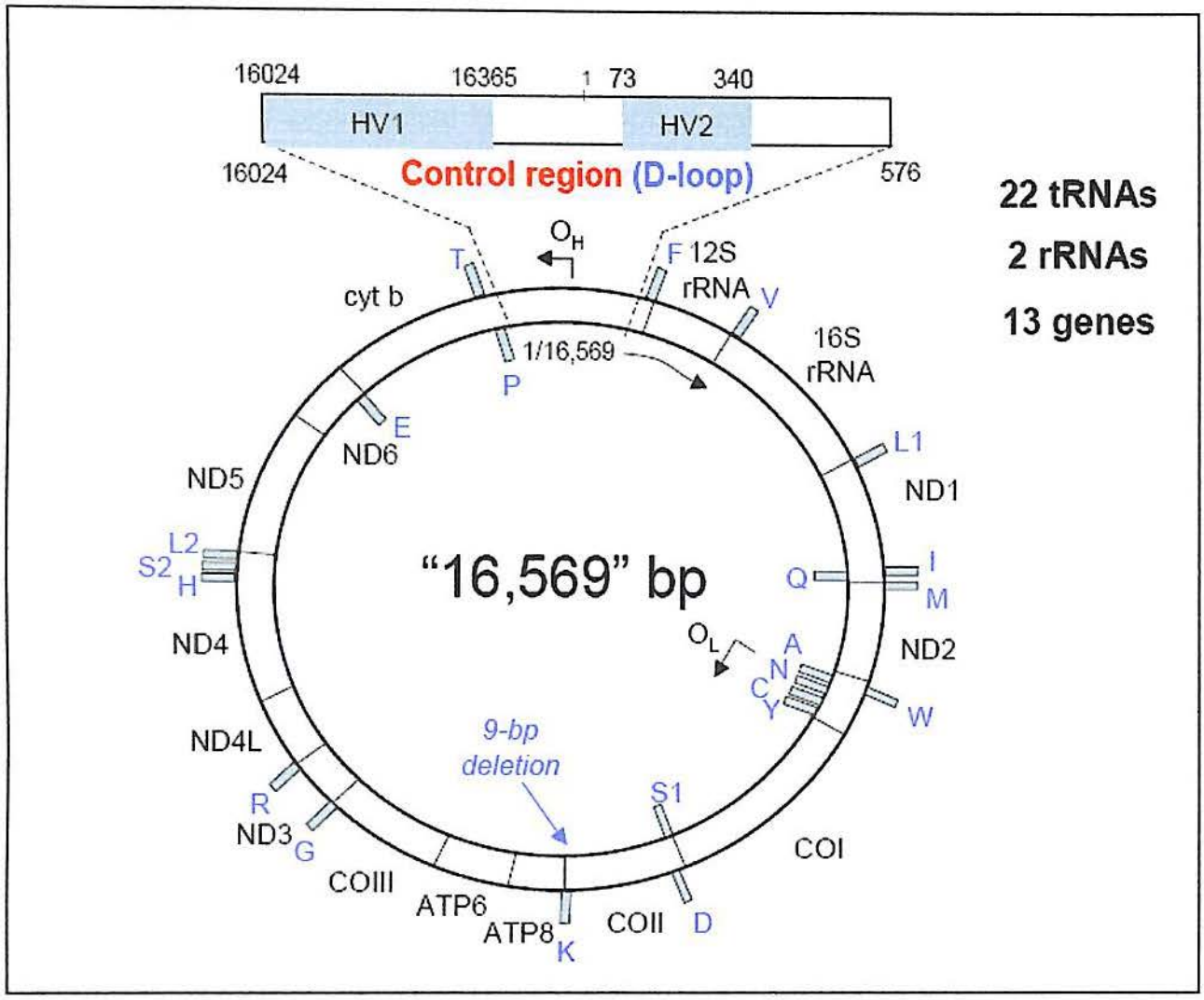

Figure 2: Organization of human mitochondrial DNA.

The human mtDNA is a 16,569 bp circular molecule that codes for seven (ND1, 2, 3, 4L, 4, 5 and 6) of the 43 subunits of complex I; one (cytochrome b, cyt b) of the 11 subunits of complex III; three (COI, II and III) of the 13 subunits of complex IV; and two (ATPase 6 and 8) of 16 subunits of complex V. It also codes for the small and large rRNAs and 22tRNAs. The heavy (H)-strand origin of replication (O_H) and the light (L)-strand origin of replication (O_L) is indicated. The D-loop is a section of mtDNA that contains the HVS-I and HVS-II regions (Butler, 2005).

FORENSIC USE OF MITOCHONDRIAL DNA TYPING

The strengths of mitochondrial DNA (mtDNA) forensic testing are well established, primarily in the three following areas: 1) successful recovery from highly degraded sources where nuclear DNA typing may fail, 2) application to cases where the only reference samples available are from matrilineal relatives, and 3) application to samples where nuclear DNA is virtually absent (e.g. shed hairs) (Holland and Parsons 1999).

As the mitochondrial DNA is arranged in a circular fashion, it is less susceptible to nucleases that tend to degrade it (Kimber, 2000). It is also afforded protection by the double-wall structure of the mitochondrion. Due to these reasons and the fact that each cell can have hundreds of copies of the genome, there is a higher probability of obtaining a successful mtDNA analysis on a sample than of a nuclear DNA analysis. Thus, samples such as skeletal remains that may have poor quantity or quality of nuclear DNA may be amenable to mtDNA typing.

Another most important forensic application of mtDNA analysis is in the analysis of hair shafts and hairs without root tissue (the type of hairs found most commonly at crime scenes). During the keratinization of hair, most of the nuclear DNA in the hair shaft is destroyed, but the mtDNA survives through this process (Linch et al., 2001).

POPULATION GENETICS STUDIES

The maternal mode of inheritance and the lack of recombination are some advantages of analyzing mtDNA for population studies (Giles *et al.,* 1980; Chen *et al.,* 1995a; Herrnstadt *et al.,* 2002; Sutovsky *et al.,* 2004). Another major advantage is its relatively high mutation rate (Stoneking, 1993), which accumulates mutations ten times faster than the nuclear genome

(Brown *et al.,* 1979; Ingman and Gyllesnsten, 2001). These properties are not found in any other locus in the human genetic makeup, thus making mtDNA the molecule of choice for the study of genetic relationships between populations.

In addition, mtDNA has been determined by Cann *et al.,* (1987) and Torroni *et al.,* (1994a) to be diverges at the rate of 2 – 4% per site per million years. This high substitution rate has been attributed to the lack of proofreading activity in mtDNA polymerase and high concentration of oxidative radicals inside mitochondria (Karmin, 2005). Among the three HVS of CR, HVS-I shows the largest variability (Lutz *et al.,* 1998; Vigilant *et al.,* 1991). In HVS-I and HVS-II, a few sites ("hotspots") exhibit very high mutation rates, and occur frequently in many different phylogenetic contexts (Wakeley, 1993; Ingman *et al.,* 2000; Stoneking, 2000). Forster *et al.,* (1996) have estimated that the mutation rate in HVS-I as about 20,180 years per transition. The fast substitution rate makes it possible to distinguish relatively recently diverged populations.

Coding region also mutate faster than genes with similar function in the nucleus (Herrnstadt *et al.,* 2002), which is about 5,140 years per base substitution (every mutation other than insertion or deletion) that has been calibrated on the basis of assumed human-chimp split 65 million years ago. The "coding region" for this calculation spans from nucleotide position (np) 577 to np 16023 (Mishmar *et al.,* 2003). It has been estimated that synonymous sites and small rRNA evolve about 20 and tRNAs about 100 times more rapidly in mitochondria than in their nuclear counterpart, which might be the result of less strict codon-anticodon pairing (Pesole *et al.,* 1999).

Rapid evolution of mtDNA has resulted in multiple RFLPs, nucleotide variants in CR and coding region (Wallace *et al.,* 1987), which are correlated with the ethnic and geographic

origin of individuals, presumably because mtDNA mutations have accumulated along with radiation of maternal lineages as women migrated out of Africa into different continents (Cann *et al.*, 1987; Stringer and Andrews 1988; Merriwether *et al.*, 1991; Torroni *et al.*, 1992; Horai *et al.*, 1995; Templeton 1998). It has been estimated that the mtDNA has undergone a mutation every 10,000 years since the first modern human, who lived approximately 150,000 – 200,000 years ago in Africa (Cann *et al.*, 1987; Vigilant *et al.*, 1991).

The mtDNA was the first polymorphic DNA structure examined in humans for evolutionary purposes (Mateu *et al.*, 1997) including prehistoric migrations and demographic events such as sudden population expansion or extreme bottlenecks (Sherry *et al.*, 1994). The first human population studies using RFLP method to detect the mtDNA variation (Denaro *et al.*, 1981; Merriwether *et al.*, 1991), revealed differences between the four great ethnic groups (Caucasian, Amerindian, African, and Asian).

MITOCHONDRIAL DNA HAPLOGROUP

Selectively neutral mutations that are well established and widely distributed among individuals of a population have been used to group mtDNA into clusters called haplogroups (Eshleman *et al.*, 2003). These are groups of mtDNAs that share a common recent ancestor who suffered a mutation that is transmitted to all of her descendants, distinguishing them from all other mtDNAs. Most of the haplogroups determining polymorphisms are continent-specific (Wallace, 1995). As a result, haplogroup allocations of a specific mtDNA lineage permit the identification of its continental origin, allowing the evaluation of matrilineal ancestry within mixed populations (Bravi *et al.*, 1997; Rando *et al.*, 1999; Green *et al.*, 2000). Clearly, measured across the passage of time, there are some mtDNAs that belong to a haplogroup that develop particular mutations. It is said that two mtDNAs that belong to the

same haplogroup but that are distinguished one from the other by additional mutations at some point of the molecule belong to distinct haplotypes (Martínez Cruzado, 2002). Haplotypes are sub-clusters of haplogroups, and the polymorphisms that determine them are less prevalent and have occurred more recently.

Haplogroups are labeled from A-Z and sub-clusters (haplotypes) are coded with a running number (Schurr *et al.*, 1990; Ballinger *et al.*, 1992; Torroni *et al.*, 1996; Richards *et al.*, 1998). The genetic relationships among haplogroups are diagramed as a genetic tree, in which the haplogroups that are on proximate branches are more closely related. Below is a diagram of the major branches and haplogroups of the human mtDNA tree (figure 3).

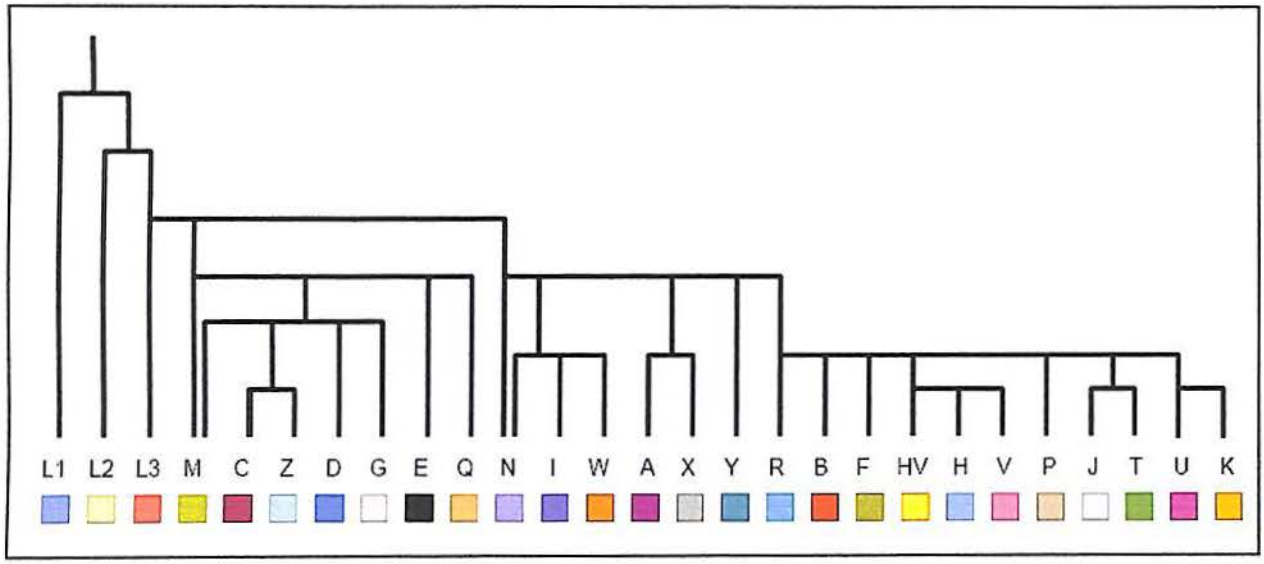

Figure 3: Simplified tree of mitochondrial haplogroups. M, N and R are called "superhaplogroups". (http://www.scs.uiuc.edu/~mcdonald/WorldHaplogroupsMaps.pdf)

MITOCHONDRIAL DNA ANALYSIS

The classification of mtDNA haplogroups is based on the variation in mtDNA which has been assessed mainly by using two different approaches: high-resolution analysis of the whole genome using restriction enzymes (restriction fragment length polymorphisms, RFLPs) (Soodyall and Jenkins, 1992; 1993; Torroni *et al.*, 1992; 2001; Chen *et al.*, 1995; 2000) or

sequencing of the HVS of CR (Vigilant *et al.,* 1991). Both approaches can be used independently, however, more comprehensive studies (Graven *et al.,* 1995; Torroni *et al.,* 1996; and Kolman *et al.,* 1996) that included the analysis of RFLPs and CR sequences for the same samples have demonstrate that the groups of mtDNA types defined by each method are correlated and both methods are complementary. Their combined use helps to clarify the phylogenetic relationship between mtDNA lineages (Richards *et al.,* 1998; Macaulay *et al.,* 1999) and led to the unification of the nomenclature by adopting the system initiated by Torroni *et al.,* (1992), which is based on RFLPs.

RFLP analysis uses a series of restriction enzymes (RE) to detect individual mtDNA variation. RE will cleave the mtDNA molecule at specific nucleotide sequences that known as recognition sites. If there are genetic differences between two individuals such that RE cut their DNA at different points in the mtDNA, the resulting fragments will be of different lengths. In RFLP analysis, the unique combination of fragments detected by a set of RE that represents genetic polymorphisms within mtDNA haplotype (Schurr, 2000).

DNA sequencing is used to survey mtDNA variation in CR. In contrast to RFLP analysis, which can be likened to a general scan of the mtDNA genome, direct sequencing provides a nucleotide-by-nucleotide reading of a portion of the mtDNA. Because the mutation rate is high in the CR, this method provides a detailed look at small genetic changes that may have taken place quite recently. Mutations in the CR help to define specific mtDNA lineages in human populations (as do certain RFLPs), and they also reveal the genetic differentiation of these lineages in geographically circumscribed areas. By statistically assessing the amount of variation in the CR within and between mtDNA lineages, it is possible to estimate the relative age of genetic variants in a particular geographic region (Schurr, 2000).

WORLD MITOCHONDRIAL DNA PHYLOGENY

Analysis of mtDNA variation has provided evidence that our species arose in Africa about 150,000 years before present (YBP) (Horai and Matsunaga, 1986; Stoneking *et al.*, 1990; Horai and Hayasaka, 1990; Di Rienzo and Wilson, 1991; Hedges *et al.*, 1991;Vigilant *et al.*, 1991), migrated out of Africa into Asia about 60,000 to 70,000 YBP and into Europe about 40,000 to 50,000YBP, and migrated from Asia and possibly Europe to the Americas about 20,000 to 30,000YBP (Wallace *et al.*, 1999). As women migrated out of Africa, additional mtDNA mutations arose and became established by genetic drift that today are seen as continent-specific mtDNA sequence polymorphisms, which known as haplogroups (Torroni and Wallace, 1994; Wallace, 1995).

Cann *et al.*, (1987) and Wallace *et al.*, (1999), supported that there was a single mtDNA tree, that the deepest root occurred in Africa, and that Africa harbored the greatest sequence diversity. Using an estimated sequence evolution rate 2-4% per million years, the human mtDNA tree was calculated to be about 2,000,000 years old (Cann *et al.*, 1987).

In West Eurasia eight haplogroups-H, I, J, T, U (including subhaplogroup K), V, W, and X-almost encompass all mtDNA variability (Torroni *et al.*, 1994; Torroni *et al.*, 1996; Macaulay *et al.*, 1999). In Asia most mtDNA types are members of superhaplogroup M (haplogroups C, D, G, and E) (Ballinger *et al.*, 1992; Torroni *et al.*, 1993; Wallace, 1995); the remaining Asian mtDNAs are encompassed by haplogroups A, B, and F (Torroni *et al.*, 1994). Among Native Americans four Asian haplogroups (A, B, C, and D) were found (Torroni *et al.*, 1993). In Africa the mtDNA lineages belong to haplogroups L1, L2, and L3 and subhaplogroup Ml (Chen *et al.*, 1995, 2000; Watson *et al.*, 1997; Quintana-Murci *et al.*, 1999).

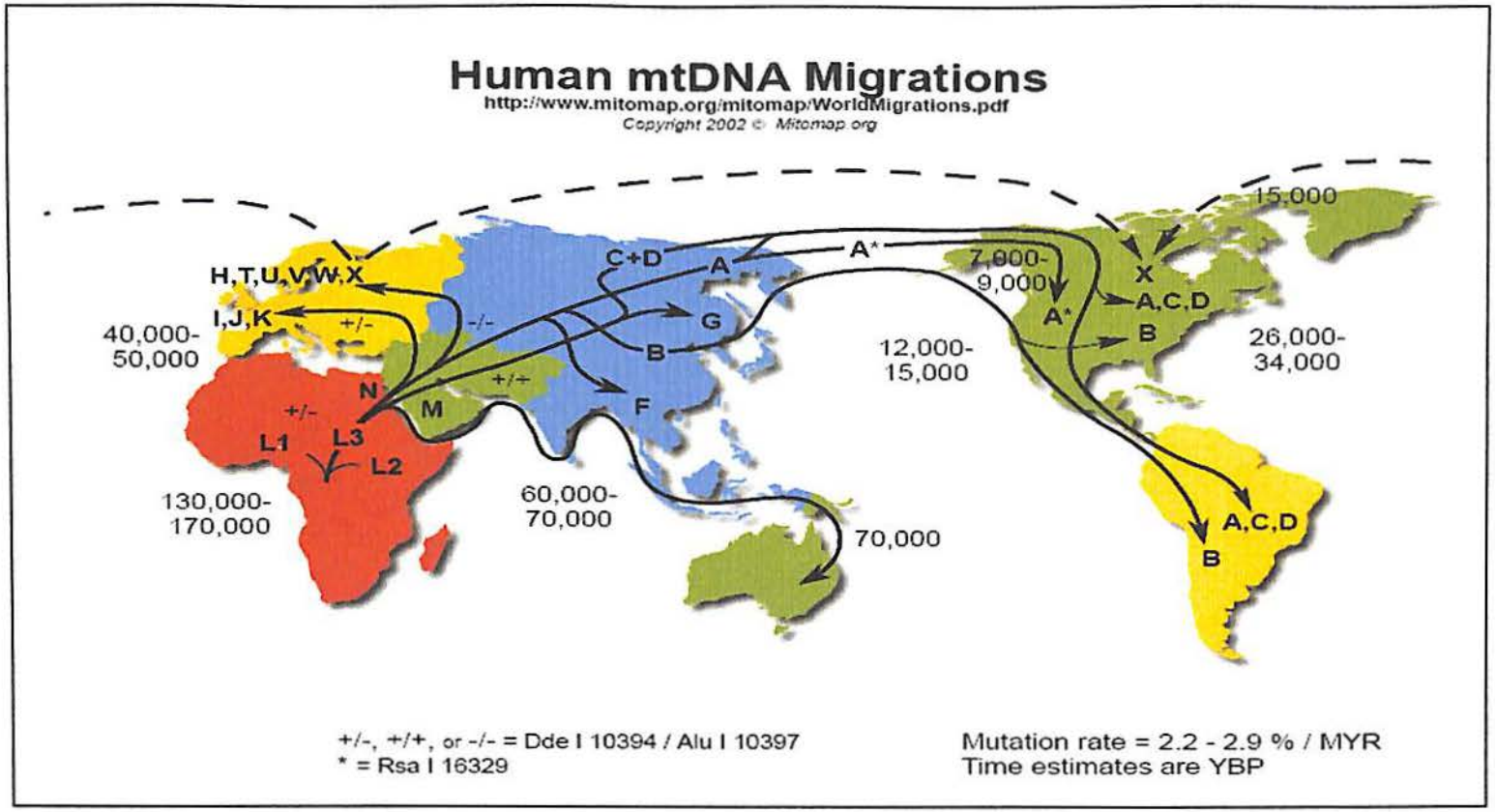

Figure 4: Human mtDNA migrations map.

Homo sapiens arose in Africa about 150,000YBP migrated out of Africa into Asia about 60,000 to 70,000YBP and into Europe about 40,000 to 50,000YBP, and migrated from Asia and possibly Europe to the Americas about 26,000 to 34,000YBP. mtDNA macro-lineage L is predominant in Africa (shown in red). mtDNA macro-lineages M and N are found throughout Eurasia and Australia. mtDNA lineages H, I, J, K, T, N, U, V, W, and X are predominant in West Eurasia. mtDNA lineages A, B, C, D, E, F, G, M are predominant in Asia and Oceania. mtDNA lineages A, B, C, D and X are found in the Americas (www.mitomap.org/WorldMigrations.pdf).

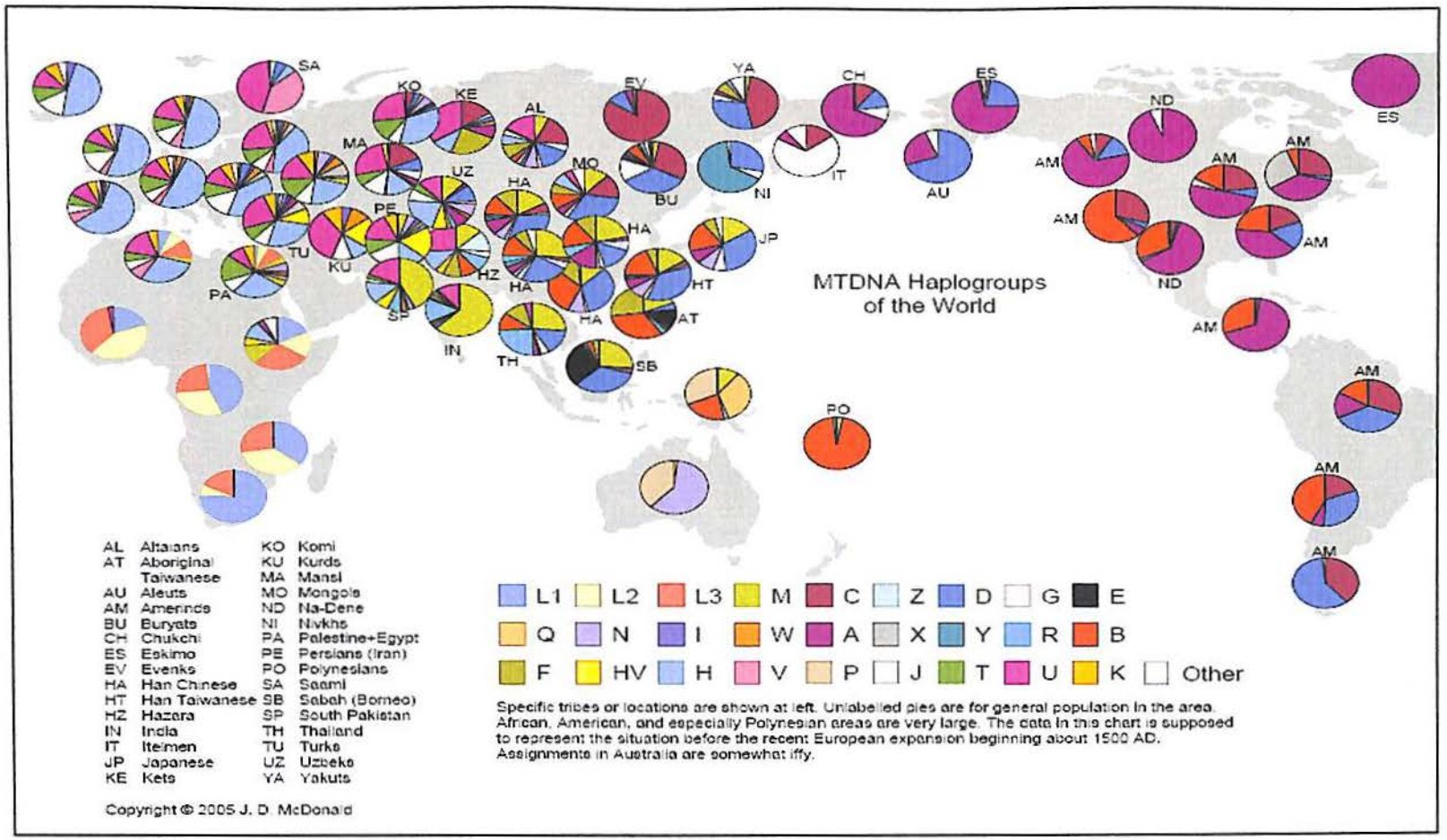

Figure 5: Distribution of the world mtDNA haplogroup.

Specific tribes or locations are shown at left. Unlabelled pies are for general population in the area. African, American, and especially Polynesian areas are very large. The data in this chart is supposed to represent the situation before the recent European expansion beginning about 1500 AD. Assignments in Australia are somewhat iffy (www.scs.uiuc.edu/~mcdonald/WorldHaplogroupsMaps.pdf).

SOUTHEAST ASIAN MITOCHONDRIAL DNA VARIATION

Among the Southeast Asian populations, several haplotypes possessed the 9-bp deletion, most of which were on a -*Dde*I (10394) and -*Alu*I (10397) background (Ballinger *et al.,* 1992). Ballinger *et al.,* (1992) have hypothesized that the 9-bp deletion arose in central China and radiated out from this region as migrants moved to populate parts of Southeast Asia. On the other hand, a significant proportion of the Southeast Asian populations possess the 9-bp 'non-deletion' allele on a +*Dde*I (10394) and +*Alu*I (10397) or +/- backgrounds (Ballinger *et al.,* 1992). Roychoudhury *et al.,* (2000) found this 9-bp 'non-deletion' allele in India.

The existence of two completely different haplotypes in Southeast Asia, with the deletion and without it, suggests that the populations of this area have two origins (Beteille, 1998; Diamond, 1997). Roychoudhury *et al.*, (2000) postulate that the early wave of migration was from India which carried the (+/+ 9-bp non-del) haplotype into Southeast Asia, occurred about 40,000 years before present (YBP). This early wave of migration from India, actually from Africa through India, to Southeast Asia has also been proposed in a recent study (Chu *et al.,* 1998) using nuclear DNA micro-satellite markers and subsequently supported by a study using Y chromosomal DNA markers (Su *et al.,* 1999). The second wave of migration from south China carried the (–/– 9-bp del) haplotype into Southeast Asia, took place about 4,000– 3,500 YBP (Ballinger *et al.,* 1992; Merriwether *et al.,* 1999).

In general, haplogroups B, F and M constitute the majority of all mtDNAs in Southeast Asian groups. One of the first haplogroups to enter Southeast Asia was clearly superhaplogroup M. Haplogroups F and B are also present throughout Southeast Asia, but are not as genetically diverse as superhaplogroup M; thus, they must have arisen after the latter

mtDNA lineage was brought into this region and began diversifying (Schurr and Wallace, 2002).

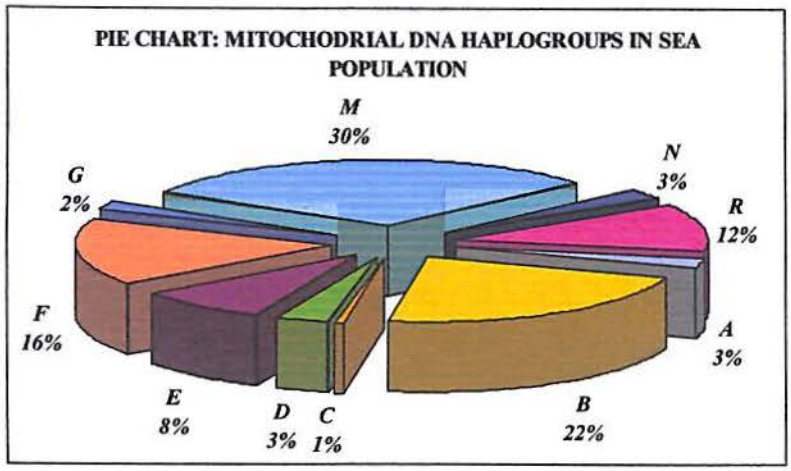

Figure 6: mtDNA haplogroup in SEA population (Ballinger *et al.,* 1992)

By contrast, haplogroups A, C, D, E, and G are less prominent in southern Asian populations, including Vietnamese, Malays, Sabah, Malay aboriginals, and New Guineans, but these groups are found at significant frequencies in Tibetans, Koreans, and Han Chinese (Torroni *et al.,* 1993b; Starikovskaya *et al.,* 1998; Schurr *et al.,* 1999).

The population tree of Southeast Asian (Ballinger *et al.,* 1992) reveals that Malays and Sabah (Borneo) Aborigines are closer to each other than to Malaysian Chinese and Vietnamese, with the Orang Asli being somewhat intermediate between the Malaysian and Vietnamese populations (Schurr and Wallace, 2002). These associations are consistent with the pattern of haplotype sharing amongst these populations (Ballinger *et al.* 1992) and with other linguistic and cultural data (e.g., Iskandar 1976; Bellwood 1979).