

**MACROHAPLOGROUP M IN MURUT  
INFERRED USING PCR-RFLP ANALYSIS OF THE  
MITOCHONDRIAL CODING REGION**

**by**

**ROHEN RAMU**

**Dissertation submitted in partial fulfillment of the requirements  
for the degree  
of Bachelor of Science (Honours) (Forensic Science)**

**January 2026**

## **CERTIFICATE**

This certifies that the dissertation titled “Macrohaplogroup M in Murut inferred using PCR-RFLP analysis of the mitochondrial coding region” is the bona fide record of research work done by Mr. Rohen Ramu during the period from October 2025 to January 2026 under my guidance and supervision. After reading this dissertation, in my opinion, it satisfies all requirements for a scholarly presentation and is fully adequate in both scope and to be submitted as a partial fulfillment for the Bachelor of Science (Honours) in Forensic Science degree.

Main supervisor,

.....

PM. Dr. Edinur Hisham Bin Atan

Lecturer

School of Health Sciences

Universiti Sains Malaysia

Health Campus

16150 Kubang Kerian

Kelantan, Malaysia

Date: .....

## DECLARATION

I hereby declare that this dissertation is the result of my own investigations, except where otherwise stated and duly acknowledged. I also declare that it has not been previously or concurrently submitted as a whole for any other degrees at Universiti Sains Malaysia or other institutions. I grant Universiti Sains Malaysia the right to use the dissertation for teaching, research and promotional purposes.

.....

(ROHEN RAMU)

Date: .....

## **ACKNOWLEDGEMENTS**

First and foremost, I would like to express my sincere gratitude to my supervisors, Assoc. Prof. Dr. Edinur Hisham Bin Atan and Dr. Nur Haslindawaty Binti Abd Rashid, for their invaluable guidance, encouragement, and continuous support throughout my final year project. I am also grateful to my laboratory supervisor for his technical assistance and for providing a supportive environment during the experimental work.

To my lab mates, Syaqirah, Azmiera, and Putri, thanks a lot for your camaraderie, support, and meaningful discussions throughout this journey. To my close group of friends, Laavanyaa, Thivya, Trisha, Kavisha, Purani, and Samuel, thank you for being there through every high and low. Your presence, laughter, and support kept me grounded throughout my undergraduate journey, and I'm truly grateful to have you all by my side.

I am profoundly grateful to my beloved family, my best friend Jay Harissh, and Parrvena, who have been constant pillars of support throughout my life. Their unconditional love, patience, and unwavering encouragement have been a source of strength and resilience, especially during the most challenging moments, and have meant more to me than words can express.

Last but not least, I wanna thank me. I wanna thank me for believing in me, I wanna thank me for doing all this hard work, I wanna thank me for having no days off, I wanna thank me for never quitting. I wanna thank me for always being a giver and trying to give more than I receive, I wanna thank me for trying to do more right than wrong. I wanna thank me for just being me at all times.

## TABLE OF CONTENTS

|                                          |             |
|------------------------------------------|-------------|
| <b>CERTIFICATE .....</b>                 | <b>II</b>   |
| <b>DECLARATION .....</b>                 | <b>III</b>  |
| <b>ACKNOWLEDGEMENTS .....</b>            | <b>IV</b>   |
| <b>TABLE OF CONTENTS .....</b>           | <b>V</b>    |
| <b>LIST OF FIGURES.....</b>              | <b>VIII</b> |
| <b>LIST OF TABLES.....</b>               | <b>IX</b>   |
| <b>LIST OF SYMBOLS .....</b>             | <b>X</b>    |
| <b>LIST OF ABBREVIATIONS .....</b>       | <b>XI</b>   |
| <b>ABSTRAK.....</b>                      | <b>XIII</b> |
| <b>ABSTRACT .....</b>                    | <b>XIII</b> |
| <b>CHAPTER 1 INTRODUCTION.....</b>       | <b>1</b>    |
| 1.1 Background of the study .....        | 1           |
| 1.2 Problem statements .....             | 3           |
| 1.3 Significance of the study .....      | 3           |
| 1.4 Objectives of the study .....        | 4           |
| 1.4.1 General objectives .....           | 4           |
| 1.4.2 Specific objectives .....          | 4           |
| <b>CHAPTER 2 LITERATURE REVIEW .....</b> | <b>5</b>    |
| 2.1 Mitochondrial genome.....            | 5           |
| 2.2 Human mtDNA haplogroups .....        | 8           |
| 2.2.1 Macrohaplogroup M .....            | 11          |
| 2.3 The Murut .....                      | 13          |
| 2.4 Application of mtDNA .....           | 16          |

|                                                 |                                                               |           |
|-------------------------------------------------|---------------------------------------------------------------|-----------|
| 2.4.1                                           | Health.....                                                   | 16        |
| 2.4.2                                           | Population genetics.....                                      | 17        |
| 2.4.3                                           | Forensic.....                                                 | 18        |
| 2.5                                             | PCR-RFLP .....                                                | 19        |
| <b>CHAPTER 3 MATERIALS AND METHODOLOGY.....</b> |                                                               | <b>21</b> |
| 3.1                                             | Materials and reagents .....                                  | 21        |
| 3.2                                             | Instruments and apparatus .....                               | 22        |
| 3.3                                             | Chemical and reagent preparation .....                        | 23        |
| 3.3.1                                           | Preparation of 10X tris-borate-EDTA TBE buffer (pH 8.3) ..... | 23        |
| 3.3.2                                           | Preparation of 1X tris-borate-EDTA TBE buffer (pH 8.3) .....  | 23        |
| 3.4                                             | Methodology.....                                              | 23        |
| 3.4.1                                           | Sample collection.....                                        | 23        |
| 3.4.2                                           | Sampling method.....                                          | 24        |
| 3.4.3                                           | Genomic DNA extraction.....                                   | 25        |
| 3.4.4                                           | Agarose gel electrophoresis of genomic DNA .....              | 26        |
| 3.4.5                                           | Genomic DNA quantification.....                               | 27        |
| 3.4.6                                           | PCR amplification of mtDNA coding region .....                | 28        |
| 3.4.7                                           | Agarose gel electrophoresis of amplified PCR products.....    | 29        |
| 3.4.8                                           | Restriction digestion .....                                   | 30        |
| 3.4.9                                           | Agarose gel electrophoresis for digested products .....       | 30        |
| <b>CHAPTER 4 RESULTS .....</b>                  |                                                               | <b>32</b> |
| 4.1                                             | Genomic DNA extraction from blood samples .....               | 32        |
| 4.2                                             | PCR.....                                                      | 34        |
| 4.3                                             | RFLP .....                                                    | 35        |
| <b>CHAPTER 5 DISCUSSION .....</b>               |                                                               | <b>42</b> |

|                                  |           |
|----------------------------------|-----------|
| <b>CHAPTER 6 CONCLUSION.....</b> | <b>45</b> |
| <b>REFERENCES .....</b>          | <b>46</b> |

## LIST OF FIGURES

|                                                                                                                                                                 | <b>Page</b> |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|
| <b>Figure 2.1:</b> Model showing the structure of the human mitochondrial genome. Adapted from Becerra (2019).                                                  | 8           |
| <b>Figure 2.2:</b> Map showing geographical distribution of human mtDNA haplogroups. Adapted from Brandon et al. (2005).                                        | 10          |
| <b>Figure 4.1:</b> Agarose gel electrophoresis image shows high molecular weight of representative extracted genomic DNA samples labelled as R052–R068.         | 32          |
| <b>Figure 4.2:</b> Representative image of amplified products (310 bp) separated using 1.2% agarose gel electrophoresis.                                        | 34          |
| <b>Figure 4.3:</b> Agarose gel electrophoresis of PCR-RFLP products following restriction digestion with (a) <i>Alu</i> I and (b) <i>Dde</i> I.                 | 36          |
| <b>Figure 4.4:</b> Proportion of Murut individuals classified as macrohaplogroup M and non-macrohaplogroup M based on PCR-RFLP analysis of mtDNA coding region. | 41          |

## LIST OF TABLES

|                                                                                                                                                            | <b>Page</b> |
|------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|
| <b>Table 2.1:</b> Distribution of major mtDNA haplogroups among several Southeast Asian populations. Data obtained from Schurr and Wallace (2002).         | 13          |
| <b>Table 3.1:</b> List of chemicals, kits, and consumables used in this research.                                                                          | 21          |
| <b>Table 3.2:</b> List of consumables used in this research.                                                                                               | 22          |
| <b>Table 3.3:</b> List of instruments and apparatus used in this study.                                                                                    | 22          |
| <b>Table 3.4:</b> Oligonucleotide primers used to amplify mtDNA coding region.                                                                             | 28          |
| <b>Table 3.5:</b> PCR components for PCR reaction mixture.                                                                                                 | 28          |
| <b>Table 3.6</b> PCR cycling parameters.                                                                                                                   | 29          |
| <b>Table 3.7</b> RFLP components for PCR mixture.                                                                                                          | 30          |
| <b>Table 4.1</b> Concentration, purity and salt carryover levels of selected extracted genomic DNA samples measured using DeNovix DS-11 spectrophotometer. | 33          |
| <b>Table 4.2</b> RFLP results showing the gain or loss of 10,397 <i>Alu</i> I and 10,394 <i>Dde</i> I restriction sites among the Murut samples.           | 37          |
| <b>Table 4.3</b> Agarose gel electrophoresis of PCR-RFLP products following restriction digestion with (a) <i>Alu</i> I and (b) <i>Dde</i> I.              | 40          |
| <b>Table 4.4</b> Classification of individuals into macrohaplogroup M based on <i>Alu</i> I and <i>Dde</i> I restriction sites.                            | 40          |
| <b>Table 4.5</b> Frequency distribution of macrohaplogroup M.                                                                                              | 41          |

## LIST OF SYMBOLS

|                      |   |                                                    |
|----------------------|---|----------------------------------------------------|
| %                    | - | Percent                                            |
| μL                   | - | Microliter                                         |
| A <sub>230</sub>     | - | Absorbance at wavelength of 230 nm                 |
| A <sub>260</sub>     | - | Absorbance at wavelength of 260 nm                 |
| A <sub>260/230</sub> | - | Absorbance ratio of wavelength at 260 nm to 230 nm |
| A <sub>260/280</sub> | - | Absorbance ratio of wavelength at 260 nm to 280 nm |
| A <sub>280</sub>     | - | Absorbance at wavelength of 280 nm                 |
| bp                   | - | Base pairs                                         |
| °C                   | - | Degree Celsius                                     |
| g                    | - | Grams                                              |
| K bp                 | - | Kilo-base pairs                                    |
| L                    | - | Liter                                              |
| M                    | - | Molar                                              |
| mL                   | - | Milliliter                                         |
| ng                   | - | Nanogram                                           |
| nm                   | - | Nanometer                                          |
| pH                   | - | Potential of hydrogen                              |
| rpm                  | - | Rotations per minute                               |
| s                    | - | Seconds                                            |
| T <sub>m</sub>       | - | Melting temperature                                |
| V                    | - | Volts                                              |

## LIST OF ABBREVIATIONS

|                      |   |                                                                            |
|----------------------|---|----------------------------------------------------------------------------|
| ATP                  | - | Adenosine triphosphate                                                     |
| CODISmt              | - | COmbined DNA index system-mitochondrial DNA                                |
| CRS                  | - | Cambridge reference sequence                                               |
| <i>cyt b</i>         | - | Cytochrome b                                                               |
| ddH <sub>2</sub> O   | - | Double distilled water                                                     |
| DNA                  | - | Deoxyribonucleic acid                                                      |
| EDTA                 | - | Ethylenediaminetetraacetic acid                                            |
| EMPOP                | - | European DNA profiling group mtDNA population database                     |
| EtBr                 | - | Ethidium Bromide                                                           |
| gDNA                 | - | Genomic deoxyribonucleic acid                                              |
| HVS-1                | - | Hypervariable segment 1                                                    |
| HVS-2                | - | Hypervariable segment 2                                                    |
| LHON                 | - | Leber hereditary optic neuropathy                                          |
| MELAS                | - | Mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes |
| MITOMAP              | - | Mitochondrial genome database                                              |
| mtDNA                | - | Mitochondrial deoxyribonucleic acid                                        |
| Na <sub>2</sub> EDTA | - | Ethylenediaminetetraacetic acid disodium salt dihydrate                    |
| NADH                 | - | Nicotinamide adenine dinucleotide                                          |
| NC                   | - | Negative control                                                           |
| ND 6                 | - | NADH dehydrogenase 6                                                       |
| OXPHOS               | - | Oxidative phosphorylation                                                  |
| PCR                  | - | Polymerase chain reaction                                                  |
| PCR-RFLP             | - | Polymerase chain reaction-restriction fragment length polymorphism         |
| rCRS                 | - | Revised Cambridge reference sequence                                       |
| RE                   | - | Restriction enzyme                                                         |
| RFLP                 | - | Restriction fragment length polymorphism                                   |

|      |   |                            |
|------|---|----------------------------|
| RNA  | - | Ribosomal nucleic acid     |
| rRNA | - | Ribosomal ribonucleic acid |
| TBE  | - | Tris-borate EDTA           |
| tRNA | - | Transfer ribonucleic acid  |
| USA  | - | United States of America   |
| UV   | - | Ultraviolet                |

## ABSTRAK

Variasi dalam genom DNA mitokondria (mtDNA) memberikan pandangan yang berharga mengenai keturunan sebelah ibu dan sejarah populasi, terutamanya dalam kalangan kumpulan peribumi yang masih kurang diwakili dalam kajian genetik. Kajian ini menganalisis taburan makrohaplogumpulan M dalam kumpulan etnik Murut menggunakan tindak balas rantai polimerase dengan analisis polimorfisme panjang serpihan sekatan (PCR-RFLP) pada kawasan pengekodan mitokondria. DNA genom diekstrak daripada sampel darah yang diperoleh daripada 100 individu Murut yang tidak mempunyai pertalian darah, diikuti dengan amplifikasi serpihan 310 bp yang mengandungi tapak polimorfik bermaklumat dalam kawasan pengekodan. Klasifikasi haplogumpulan ditentukan melalui pencernaan sekatan menggunakan enzim *Alu I* dan *Dde I* yang menyasarkan kedudukan nukleotida 10,397 dan 10,394 masing-masing. Corak serpihan PCR sekatan membolehkan pembezaan yang boleh dipercayai antara keturunan makrohaplogumpulan M dan bukan M. Makrohaplogumpulan M didapati kurang kerap (0.25) tersebar dalam etnik Murut berbanding populasi lain yang berkaitan dari segi keturunan dan bahasa seperti orang Melayu (0.57). Corak penempatan tradisional kaum Murut, aliran gen yang terhad dengan populasi pesisir pantai atau migran, serta saiz populasi berkesan yang agak kecil mungkin menyumbang kepada kekerapan makrohaplogumpulan yang lebih rendah dalam kalangan peribumi Sabah ini. Secara keseluruhannya, penemuan ini menyumbang data asas yang berharga mengenai struktur genetik ibu kumpulan etnik Murut dan meningkatkan pemahaman semasa mengenai kepelbagaian mtDNA dalam kalangan populasi peribumi di Asia Tenggara.

## ABSTRACT

Variations in mitochondrial DNA (mtDNA) genome provides valuable insight into maternal ancestry and population history, particularly among indigenous groups that remain underrepresented in genetic studies. This study analyzed the distribution of mtDNA macrohaplogroup M in the Murut ethnic group using polymerase chain reaction with restriction fragment length polymorphism (PCR-RFLP) analysis of the mitochondrial coding region. Genomic DNA was extracted from blood samples obtained from 100 unrelated Murut individuals, followed by amplification of a 310 bp fragment containing informative polymorphic sites within the coding region. Haplogroup classification was determined through restriction digestion using *Alu* I and *Dde* I enzymes targeting nucleotide positions 10,397 and 10,394, respectively. The restriction PCR fragment patterns enabled reliable differentiation between macrohaplogroup M and non-M lineages. The macrohaplogroup M is less frequently (0.25) distributed in Murut as compared with other ancestrally and linguistically related populations such as Malays (0.57). The Murut's traditional settlement patterns, limit gene flow with coastal or migrant populations, and relatively small effective population size may have contributed to the less frequent of macrohaplogroup in this native of Sabah. Overall, the findings contribute the valuable baseline data on the maternal genetic structure of the Murut ethnic group and enhance current understanding of mtDNA diversity with indigenous populations of Southeast Asia. Future studies should include other natives of Sabah and Sarawak (e.g., Bajau, Bugis, Kedayan, Iban and Kadazan) as their macrohaplogroup M frequency distributors have yet to be determined.

## CHAPTER 1

### INTRODUCTION

#### 1.1 Background of the study

Mitochondrial DNA (mtDNA) polymorphism analysis serves as a useful tool in human population genetics, enabling the rigorous reconstruction of ancestral maternal migration routes and the demographic history of human dispersal across continents (Malyarchuk et al., 2002). This approach is predicated on the unique characteristics of the mitochondrial genome, a highly conserved, circular, double-stranded molecule approximately 16,569 base pairs in length that is inherited exclusively through the maternal lineage without recombination (Naini et al., 2020). The absence of recombination and the high copy number (hundreds to thousands per cell) render mtDNA an indispensable marker for tracing evolutionary relationships over deep timescales, especially for highly degraded ancient human specimen (Cavalcanti et al., 2024). The human mitochondrial genome encodes 37 genes which are important for transcription and translation of proteins for oxidative phosphorylation system (Habbane et al., 2021).

The focus of this study is Macrohaplogroup M, one of the two major founding macrohaplogroups (other than N) that arose from the African lineage L3 and encompasses virtually in all non-African maternal diversity globally (Marrero et al., 2016). Macrohaplogroup M exhibits extensive geographical dominance across Asia, frequently comprising more than 70% of maternal lineages across the Indian subcontinent, East Asia, and Southeast Asia (Chandrasekar et al., 2013). The dispersal history of the macrohaplogroup M remains subject to evolutionary debate, with theories proposing its

origin in South Asia or Southeast Asia. Several molecular analyses suggesting that the colonization waves carrying M lineages often originated from the east (Southeast Asia/Oceania) to settle regions further west, positioning the island regions of Malaysia as a long-term reservoir of this ancient genetic heritage (Marrero et al., 2016). The ubiquitous presence and deep diversity of M in Southeast Asia, including its specific subclades such as Haplogroup E, emphasize the region's crucial role in macrohaplogroup M's ancient dispersal history.

This study involves macrohaplogroup M characterization in one of the distinct indigenous groups residing in Borneo, Malaysia, positioned centrally within this critical dispersal corridor, the Murut population. The Murut, traditionally known as "hill people" or *Ulun Tahol* (Tahol people), typically inhabit the interior highlands of Sabah and regions like the Tomani area (Mohad et al., 2016). Analyzing macrohaplogroup M in this community is pivotal for elucidating the ancient maternal migratory events and genetic connections that tie Murut to other populations in the broader Asian Pacific region.

This study involves polymerase chain reaction (PCR) of the mtDNA coding region and followed by amplicons digestion with restriction enzymes (RFLP). The PCR-RFLP is a cost-effective and sensitive genotyping technique designed to detect single nucleotide variations by analyzing the different lengths of DNA fragments digested with restriction enzymes (Tarach, 2021). The analysis targets the mtDNA coding region, spanning approximately 15.5 kilobase pairs. This detailed approach is mandated for accurately classifying mitochondrial lineages because analyzing haplogroup-specific coding region polymorphisms provides the resolution necessary to circumvent the ambiguity caused by parallel mutations (homoplasy) frequently observed when relying solely on the highly

variable non-coding region of mtDNA genome (van Oven & Kayser, 2009). Thus, the methodology employed offers a robust and established methodological approach for precisely characterizing the unique maternal macrohaplogroup M genetic heritage in Murut.

## **1.2 Problem Statements**

The mtDNA has long been used as a genetic marker for studying evolutionary history and population relationships (Pakendorf & Stoneking, 2005). In Malaysia, several studies have examined the maternal genetic structure of indigenous populations such as the Murut group through analysis of the mtDNA control region (Lian & Koh, 2005). However, relying solely on the control region can lead to incomplete or ambiguous haplogroup classification as this region prone to homoplasmy (Nilsson et al., 2007).

Thus, an accurate and reliable determination of mtDNA haplogroups rely on coding region, which contains stable polymorphic sites that serve as diagnostic markers for specific macrohaplogroup determinations. Despite its importance, no study has yet focused on coding region of mtDNA to identify macrohaplogroup M frequency distributions in Murut.

## **1.3 Significance of the study**

This study enables direct detection of restriction site gains or losses caused by specific nucleotide substitutions that define macrohaplogroup M via PCR-RFLP technique. Thus, this study will provide a better insight on mtDNA types in Murut and compliment earlier mtDNA study (Lian & Koh, 2005) in this Borneo population group. Moreover, this study provides mtDNA data in Borneo and Asia Pacific region. In essence,

this study's innovative approach of shifting from control region analysis to coding region confirmation enhances scientific accuracy, strengthens methodological robustness, and provides a more conclusive insight into the maternal ancestry of the Murut community.

## **1.4 Objectives of the study**

### **1.4.1 General objectives**

To identify mtDNA macrohaplogroup M in Murut ethnic group using PCR-RFLP technique.

### **1.4.2 Specific objectives**

- i. To amplify mtDNA coding region in DNA samples collected from unadmixed and unrelated Murut individual using PCR technique.
- ii. To digest the PCR amplified products using *Alu* I and *Dde* I.
- iii. To determine the macrohaplogroup M frequency in Murut based on the *Alu* I and *Dde* I digested patterns.

## CHAPTER 2

### LITERATURE REVIEW

#### 2.1 Mitochondrial genome

Mitochondria are essential double-membrane organelles located within the cytoplasm of eukaryotic cells, often described as the powerhouse of the cell. Their primary function is to transform energy in the form of adenosine triphosphate (ATP) from food via series of oxidative phosphorylation process (Habbane et al., 2021). Mitochondria are unique because they contain their own independently maintained genome known as mitochondrial DNA (mtDNA), which is separate and distinct from the other type of genome located in the cell nucleus (nuclear genome).

Genetic elements in mitochondria were first identified and isolated by Margit Nass and Sylvan Nass in 1963 (Amorim, Fernandes, & Taveira, 2019). Nearly two decades later, Anderson and colleagues published the first complete sequence of the human mtDNA genome, widely known as the Cambridge Reference Sequence (CRS) (Anderson et al., 1981). Subsequent refinements by Andrews et al. (1999) led to the establishment of the Revised Cambridge Reference Sequence (rCRS), which now serves as the universal benchmark for reporting and comparing mitochondrial sequence variations.

The human mitochondrial genome is a circular and double-stranded molecule that is approximately 16,569 base pairs (bp) in length and unlike the nuclear genome, it has no introns. This genome is composed of two distinct strands; a purine-rich strand and a pyrimidine-rich strand which are well differentiated by their nucleotide composition and

density. The guanine-rich strand is termed the heavy (H) strand, while the cytosine-rich strand is known as the light (L) strand. The H strand encodes most of the genetic information, including 12 of the 13 polypeptide-encoding genes, 14 of the 22 tRNA-encoding genes, and two rRNA-encoding genes. The L strand encodes one polypeptide (or a single protein-coding gene) and eight tRNAs (Amorim, Fernandes, & Taveira, 2019).

MtDNA exists in hundreds to thousands of copies per cell, with its specific abundance varying according to a tissue's energetic demands. This high copy number contrasts with nuclear DNA, which is limited to two copies per cell (Pakendorf & Stoneking, 2005). Another unique characteristic of mtDNA is its maternally inherited pattern and lack of genetic recombination. Thus, mtDNA is passed as a single unit or haplotype from a mother to all her children (Budowle et al., 2000).

The mitochondrial genome is divided into two; the coding region and the control region (Figure 2.1). The coding region contains 13 protein-coding genes associated with oxidative phosphorylation (OXPHOS) system, 22 tRNAs required for intra-mitochondrial protein synthesis, and two rRNAs essential for the mitochondrial ribosome (Pakendorf & Stoneking, 2005). The control region, also known as the D-loop, is a small non-coding segment of about 1,100 bp that regulates how the genome is copied and used. This region includes hypervariable segments such as hypervariable segment (HVS)-1, which are the most polymorphic region in mtDNA genome and widely used for population genetic analyses and individual identification (Budowle et al., 2000).

The key characteristic of mtDNA include a high mutation rate, which is estimated to be 10 to 20 times faster than that of nuclear DNA. This high mutation rate is caused by replication errors and its proximity to reactive oxygen species produced during energy generation (Budowle et al., 2000). Since there are so many copies of mtDNA, an individual can have a mixture of both healthy and mutated DNA, a state called heteroplasmy. Clinical symptoms associated with mtDNA typically only appear when the amount of mutated DNA passes a certain threshold, often between 60% and 80% (Naini et al., 2020). During cell division, these different types of mtDNA are distributed randomly to daughter cells in a process called mitotic segregation, which can cause the level of mutation to vary between different tissues in the same person (Naini et al., 2020).

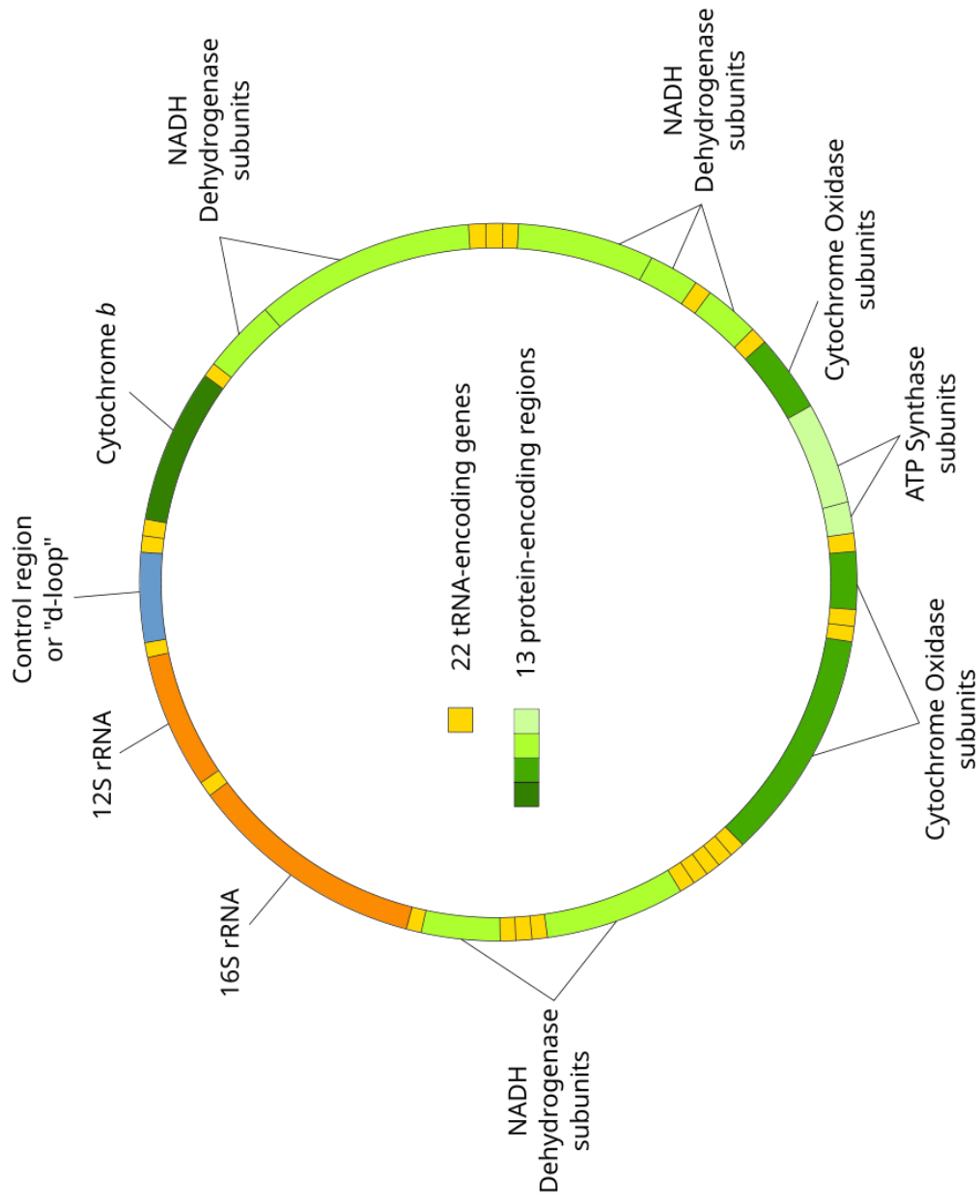


Figure 2.1: Model showing the structure of the human mitochondrial genome. Adapted from Becerra (2019).

## **2.2 Human mtDNA haplogroups**

Human mtDNA evolves through the sequential accumulation of mutations along maternally inherited lineages (van Oven & Kayser, 2009). These well-established, neutral mutations allow researchers to categorize individuals into distinct clusters called haplogroups, which represent groups of people who share a common maternal ancestor (Chuen, 2007). Globally, all human mitochondrial lineages originate from a common ancestor in Africa, with all non-African lineages descending from major macrohaplogroup L3 (Fernandes et al., 2012).

Emerging approximately 60,000 to 70,000 years ago, macrohaplogroup L3 splitted into two primary non-African nodes; macrohaplogroups M and N (Marrero et al., 2016). Macrohaplogroup M is widely distributed throughout Asia and Oceania, encompassing specific subclades such as C, D, G, and Z, while macrohaplogroup N (and its major subclade R) includes the ancestral lineages for European populations (e.g., H, I, J, T, and U) as well as various Asian and Australasian groups like A, B, and F macrohaplogroups (van Oven & Kayser, 2009). Within these broad categories, more specific haplogroups are defined by more recent mutations, offering a highly resolved map of individual maternal heritage (Chuen, 2007). The global human migration patterns inferred from mtDNA variation and continent-specific mtDNA haplogroup distributions are shown in Figure 2.2.

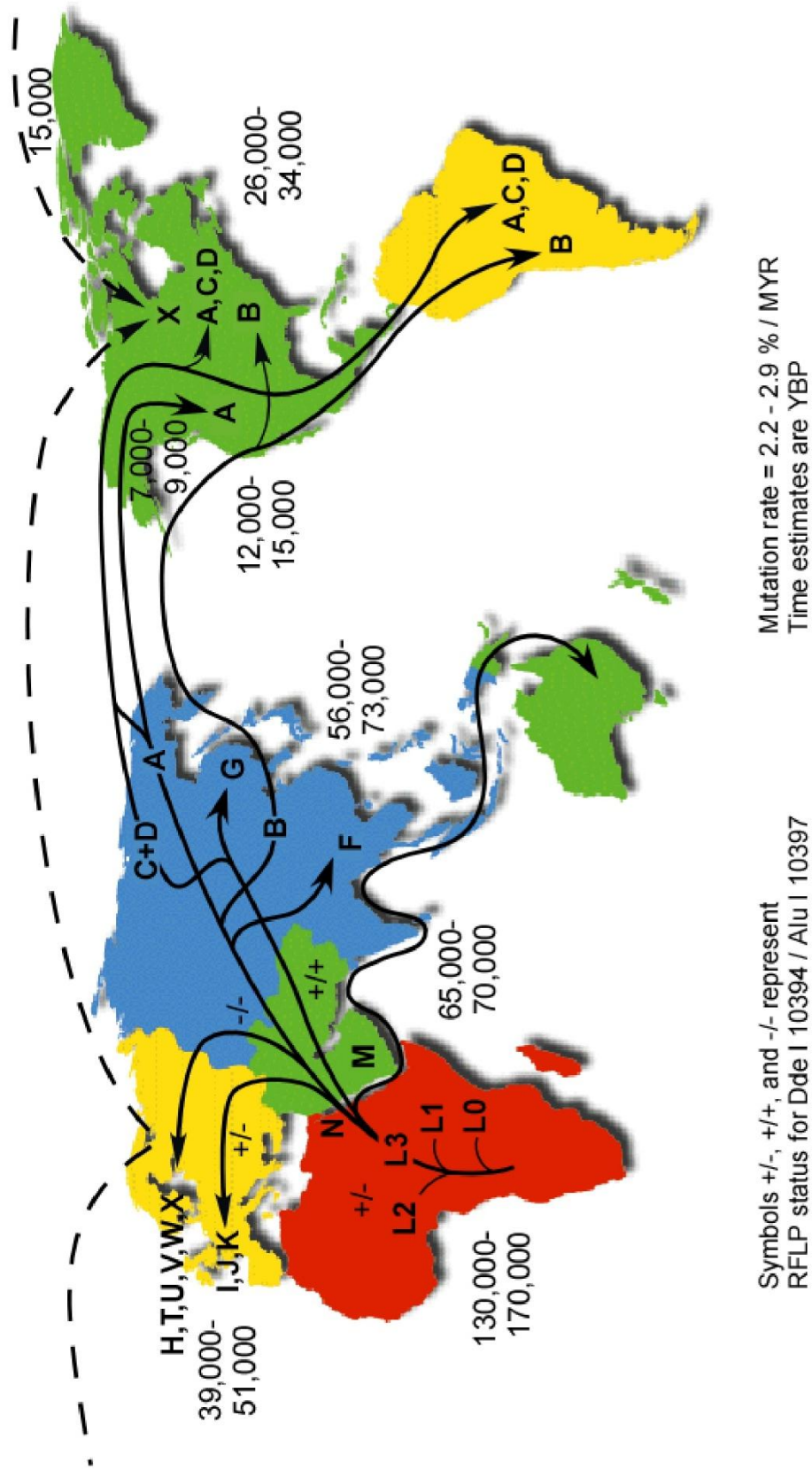


Figure 1.2: Map showing geographical distribution of human mtDNA haplogroups. Time estimates are expressed thousands of years ago (ya) and are based on a mutation rate of 2.2–2.9% per million years. The symbols +/+, +/-, and -/- indicate the Dde I 10394/ Alu I 10397 polymorphisms, while \* denotes the presence of the +Rsa I site. Adapted from Brandon et al. (2005)

### 2.2.1 Macrohaplogroup M

Human macrohaplogroup M, alongside macrohaplogroup N, is one of the two primary mtDNA lineages that diverged from the African L3 ancestor approximately 60,000 to 70,000 years ago. Macrohaplogroup M is defined by the presence of an *Alu* I restriction site at nucleotide position 10,397, which results from a C-to-T transition at nucleotide position 10,400, and *Dde* I restriction site at nucleotide position 10,394, created by an A-to-G transition at nucleotide position 10,398 (Zainuddin & Ek, 2004). The absence of the nucleotide position 10,397 *Alu* I restriction site, and in some cases the nucleotide position 10,394 *Dde* I site, indicates that the mitochondrial DNA belongs to macrohaplogroup N (Lell & Wallace, 2000).

The macrohaplogroup M is widely distributed throughout Asia and Oceania and only introduced in Europe via recent historical migrations (González et al., 2007). In East Asia, macrohaplogroup M includes prominent clades such as C, D, G, and Z, whereas South Asian populations harbor unique, deep-rooting lineages like M2 and M6 (Metspalu, 2004). Recent phylogeographic analyses suggest that macrohaplogroup M followed a complex dispersal route, with evidence indicating that macrohaplogroup M lineages in India are significantly younger than those found in East Asia, Southeast Asia, and Near Oceania (Marrero, Abu-Amero et al., 2016).

This age gradient suggests that macrohaplogroup M-carrying populations may have colonized the Indian subcontinent from the east (Southeastern Asia) rather than from the west, as was previously assumed under the "southern coastal route" model (Marrero et al., 2016). Furthermore, haplogroup M1 is recognized as the only variant of the M

lineage present in Africa, which is believed to trace an early human backflow from Asia into Northern and Eastern Africa during the Late Pleistocene (González et al., 2007).

In specific regional contexts, macrohaplogroup M remains as the most frequently distributed mtDNA type and accounts for approximately 55% of the maternal gene pool in the Chinese population of Peninsular Malaysia, where it is frequently categorized into sub-clusters MC, MD, and MG (Chuen, 2007). Other indigenous groups within the region show distinct macrohaplogroup M distributions, 39.4% for Orang Asli and 56.3% for aborigines of Sabah. In the broader East Asian and Oceanic contexts, macrohaplogroup M is found at 35% in Taiwanese, 32.1% in Vietnamese, and 44.5% in Papua New Guinea populations (Table 2.1) (Schurr & Wallace, 2002).

While frequencies remain high in neighboring regions such as Korea (61.5%) and Tibet (61.1%) (Schurr & Wallace, 2002), a significant maternal genetic boundary exists to the west, where frequencies drop abruptly from over 65% in India to approximately 5% in Iran, marking a significant maternal genetic boundary (Marrero et al., 2016). Since these lineages consist of selectively neutral mutations established over thousands of years, they serve as a powerful tool for tracing prehistoric human migrations and defining the geographic origins of maternal ancestry.

Table 2.1: Distribution of major mtDNA haplogroups among several Southeast Asian populations. Data obtained from Schurr and Wallace (2002).

| <b>Haplogroup</b><br><b>Population</b> | n   | B    | F    | M    | Other |
|----------------------------------------|-----|------|------|------|-------|
| <b>Taiwanese</b>                       | 20  | 7.1  | 22.2 | 35.0 | 35.7  |
| <b>Malaysian Han</b>                   | 14  | 32.0 | -    | 35.7 | 32.3  |
| <b>Vietnamese</b>                      | 28  | 17.9 | 32.1 | 32.1 | 17.9  |
| <b>Orang Asli</b>                      | 31  | 3.1  | 27.3 | 39.4 | 30.2  |
| <b>Malays</b>                          | 14  | 14.3 | 14.3 | 57.1 | 14.3  |
| <b>Sabah Aborigines</b>                | 32  | 18.6 | 2.9  | 56.3 | 22.2  |
| <b>Papua New Guinea</b>                | 119 | 41.8 | 1.0  | 44.5 | 12.7  |
| <b>Koreans</b>                         | 13  | 15.4 | 7.7  | 61.5 | 15.4  |
| <b>Tibetans</b>                        | 54  | 11.1 | 14.8 | 61.1 | 13.0  |
| <b>Nepalese</b>                        | 50  | 4.0  | 4.0  | 30.0 | 62.0  |

Note: "Other" encompasses all mtDNA that do not fall into macrohaplogroups B, F, and M.

### 2.3 The Murut

The Murut people represent a significant indigenous community located deep within the rugged landscapes of Borneo and are particularly concentrated in Sabah. The term "Murut" literally translates to 'hill people', reflecting their traditional habitation in highland areas (Joshua, 2025). The Murut are geographically dispersed across various parts of Borneo Island, encompassing Brunei, Kalimantan (Indonesia), and the East Malaysian states of Sabah and Sarawak. Their largest population center is in Sabah (Mohad et al., 2016).

They also inhabit the rural Temburong district in Brunei. Geographically, their territory traditionally spanned the interior from Brunei Bay toward the headwaters of the Padas River, extending to the Baram headwaters, and into the Indonesian mountains of East Kalimantan, where tributaries of the Sesayap River originate (Joshua, 2025). The Murut are composed of several sub-groups, each possessing its own distinct dialect. The largest subgroups identified include Tagal, Tidung, Timugon, Sembakung, Paluan, Bookan, Kalabakan, and Serundung Murut (Mohad et al., 2016).

In terms of population size, the Sabah Murut community is estimated to be around 135,000, with approximately 1,200 found in Brunei (Harvest, 2025). In context of their language, Murut speaks several related languages, collectively known as the Murutic languages, which are part of the Austronesian language family. Each subgroup has their own dialect, but most Murut are also conversant in Malay, the national language of Brunei and Malaysia (Joshua, 2025).

The history of the Murut is marked by their fierce warrior traditions. They are descendants of Borneo's legendary headhunters and were among the last tribal groups in Borneo to abandon the practice of headhunting, a custom interwoven with tribal warfare and spiritual meaning. Despite the cessation of this ancient practice, the spirit of the Murut warrior is still invoked in their commitment to defending their land, culture, and way of life (Mohad et al., 2016).

Traditionally, the Murut were shifting cultivators, moving their settlements every few years. Their traditional settlements were communal longhouses, typically constructed near rivers. However, this lifestyle has shifted, and they have since abandoned the

longhouse living style for individual family houses, although these modern villages are generally located in the areas of their former communities (Joshua, 2025). Their livelihood was historically based on hunting, subsistence farming, and fishing. They cultivated staple crops such as tapioca and hill rice. Murut men demonstrated great skill as hunters, employing blowpipes, spears, and hunting dogs (Joshua, 2025).

Currently, cultivating hill rice remains as primary income generation for Murut. Other avenues of employment include careers in the military, saw milling, and timber processing. In Brunei, the Murut population has generally benefitted from the nation's modernization and economic prosperity. Similarly, the Murut in Sabah have seen increased opportunities due to modernization, though individuals living in remote locations have not benefitted as extensively from these modernization (Mohad et al., 2016).

Christianity is the predominant religion for Murut in both Brunei and Sabah. However, this religion designation often used culturally to distinguish their contemporary way of life from their previous culture and from the predominant Muslim culture, rather than necessarily describing individual religious conviction (Joshua, 2025). The Murut are often recognized as the "guardians of Sabah's soul" (Tours, 2024).

## **2.4 Application of mtDNA**

### **2.4.1 Health**

MtDNA is essential for human physiology because it encodes components of cellular energy production. Mutations in mtDNA can disrupt mitochondrial function and have been linked to a variety of diseases, particularly age-related and degenerative disorders (Wallace, 1994; Naini et al., 2020). Such mutations are commonly studied in energy-dependent tissues, with associations reported in mitochondrial cardiomyopathies and neurodegenerative diseases such as Parkinson's disease and Alzheimer's disease (Wei & Chinnery, 2020).

A critical aspect of mtDNA genetics is heteroplasmy, a condition where a mixture of both healthy and mutated DNA exists within a cell. Clinical symptoms typically only appear once the level of mutated DNA passes a specific pathogenic threshold, often estimated to be between 60% and 80% (Naini et al., 2020). This threshold effect explains why tissues with high energy needs, such as the brain, heart, and skeletal muscles, are particularly vulnerable to these mutations. Well-documented examples include the m.3243A>G mutation in mitochondrial encephalomyopathy, lactic acidosis and stroke-like episodes syndrome (MELAS syndrome) and the m.11778G>A mutation in Leber hereditary optic neuropathy (LHON) (Habbane, et al., 2021).

### **2.4.2 Population genetics**

The high mutation rate of mtDNA, allow accurate reconstruction of maternal lineages and trace human evolutionary history back through time without the confounding effects of biparental inheritance (Cavalcanti et al., 2024). While the entire genome is informative, the non-coding control region, particularly the HVS-1 and HVS-2, accumulates mutations approximately 10 times faster than the coding region, providing a high degree of variation among individuals (Pakendorf & Stoneking, 2005).

Sequence variations in mtDNA is used to classify individuals into haplogroups, which are defined by shared mutations and often show strong geographic associations (Cavalcanti et al., 2024). Early studies commonly employed RFLP analysis of the control region, particularly the HVS-1 and HVS-2, to investigate human migration patterns, population divergence, and settlement history (Malyarchuk et al., 2002). Such approaches have been instrumental in studies of global human dispersal, including the colonization of the Pacific region (Friedlaender et al., 2007). However, since mtDNA reflects only maternal ancestry, it is often complemented with Y-chromosome and autosomal markers to provide a more comprehensive view of population history and to account for potential biases such as genetic drift or selection (Pakendorf & Stoneking, 2005).

### **2.4.3 Forensic**

Information obtained from mtDNA analysis is crucial in forensic genetics, particularly when nuclear DNA (nDNA) is too degraded or present in exiguous quantities (Cavalcanti et al., 2024). Its unique circular structure and high copy number, often reaching hundreds to thousands of copies per cell, enable successful profiling from challenging specimens such as hair shafts, old bones, teeth, and charred remains (Pakendorf & Stoneking, 2005). These characteristics make mtDNA especially valuable for missing person investigations and for identifying mass disaster victims. The latter include victim identification of 2004 Southeast Asia tsunami (Deng et al., 2005). Since mtDNA is inherited exclusively from the mother and lacks genetic recombination, maternal relatives can be utilized as reference sources when direct reference samples are unavailable for comparison (Cavalcanti et al., 2024).

While earlier forensic efforts primarily focused on sequencing of HVS-1 and HVS-2 of the control region, this limited scope often lacked the discriminatory power required for unique identification (Li et al., 2025). Modern research indicates that for certain haplogroups, over 70% of variations exist outside these segments, prompting a significant shift toward complete mitochondrial genome sequencing to improve individual differentiation (Li et al., 2025). To ensure standardized and reliable results, these profiles are interpreted using international guidelines and validated against global mtDNA population databases like EMPOP, CODISmt, and MITOMAP, which significantly strengthen the statistical robustness of forensic casework (Cavalcanti et al., 2024).

## 2.5 PCR-RFLP

The PCR-RFLP is one of the molecular techniques frequently used to study sequence variations within the mtDNA genome (Pakendorf & Stoneking, 2005). Initially, studies of human mtDNA variation relied upon restriction digestions of either purified genomic DNA or purified mtDNA. This technique became highly influential as RFLP analysis of the mtDNA coding region successfully revealed a number of stable polymorphic sites, which proved essential for classifying mtDNA haplogroups (Zainuddin & Ek, 2004). Following the innovation of the polymerase chain reaction (PCR), RFLP analysis was adapted to examine PCR products, often focusing on fragments from both the coding and non-coding regions for mtDNA haplogroup characterization.

The PCR-RFLP technique requires several deliberate steps to identify polymorphisms in mtDNA. Initially, specialized primers are used to amplify fragments of the mtDNA coding region that contain the targeted polymorphic sites (Zainuddin & Ek, 2004). These fragments can range in size; for instance, one study utilized a lengthy 2,790 bp PCR product encompassing genes like *cyt b* and *ND 6*, as well as a section of the D-loop (Lee et al., 2006). The amplified products are then subjected to restriction digestion using specific restriction enzymes (REs). These enzymes are selected based on criteria such as being low priced, readily available, and active in the PCR mix, while maximizing the differentiation potential (Juma, 2024). Depending on the desired resolution, analysis ranges from low resolution (use up from five to six restriction enzymes) and up to high resolution (use up from 12 to 14 enzymes) (Chen et al., 2000; Cann et al., 1987; Macaulay et al., 1999; Zainuddin & Ek, 2004). The most common set of restriction endonucleases used in the RFLP analysis includes *Alu I*, *Ava II*, *BamHf I*, *Dde I*, *Hae II*, *Hae III*, *Hha I*,

*Hinc* II, *Hinf* I, *Hpa* I, *Mse* I, *Msp* I, *Mbo* I and *Rsa* I. These restriction enzymes cleave the DNA at specific recognition sites, and the resulting fragments are separated and visualized using agarose or polyacrylamide gel electrophoresis, typically stained with ethidium bromide (Zainuddin & Ek, 2004).

The interpretation of RFLP analysis relies on observing changes in the fragment profiles following electrophoresis, where site differences among haplotypes are inferred based on the gain or loss of particular restriction sites (Chaudhary & Maurya, 2019; Juma, 2024; Zainuddin & Ek, 2004). If a single nucleotide substitution creates a new restriction site, the enzyme cuts the fragment, reducing its size. Conversely, if a mutation causes the loss of a naturally occurring restriction site, the fragment remains intact or appears longer (Juma, 2024). The final outcome, consisting of distinct restriction morphs, is documented by assigning a multi-letter code (haplotype) to each specimen (Lee et al., 2006).

This method proved vital for mtDNA classification. For example, specific haplogroups like M are defined by the presence of an *Alu* I site at position 10,397 and a *Dde* I site at 10,394 (Zainuddin & Ek, 2004). Despite its utility, analysis using a common set of enzymes examines only about 20% of the entire mtDNA genome, and RFLP results can present ambiguities regarding the actual number of nucleotide differences compared to direct sequencing (Wallace, 1994; Zainuddin & Ek, 2004). Phylogenetic and forensic studies often compensate by using RFLP data in combination with sequence variation analysis from the hypervariable regions for a more precise classification of mtDNA genotypes (Macaulay et al., 1999; Malyarchuk et al., 2002; Zainuddin & Ek, 2004; Torroni et al., 1996).

## CHAPTER 3

### METHODOLOGY

#### 3.1 Materials and reagents

All the chemicals, reagents, kits, and consumables used in this study and their respective providers are listed in Table 3.1 and Table 3.2.

Table 3.1: List of chemicals, kits, and consumables used in this research.

| Chemical                                                              | Production Company                   |
|-----------------------------------------------------------------------|--------------------------------------|
| 100bp ladder                                                          | ThermoFisher Scientific, USA         |
| 1kbp DNA ladder                                                       | Gene DireX, Taiwan                   |
| Agarose powder                                                        | 1 <sup>st</sup> Base, Singapore      |
| Boric acid                                                            | Promega Corporation, USA             |
| Disodium ethylenediaminetetraacetate dehydrate (Na <sub>2</sub> EDTA) | Sigma-Aldrich, USA                   |
| Ethidium bromide                                                      | Sigma-Aldrich, USA                   |
| Orange G Dye                                                          | Sigma-Aldrich, USA                   |
| Primers                                                               | Integrated DNA Technology, Singapore |
| PrimeWay Genomic II DNA Extraction Kit                                | 1 <sup>st</sup> Base, Singapore      |
| Restriction Enzymes                                                   | Vivantis Technologies, Malaysia      |
| Tris base                                                             | Sigma-Aldrich, USA                   |

Table 1.2: List of consumables used in this research.

| <b>Consumables</b>     | <b>Production Company</b>           |
|------------------------|-------------------------------------|
| 0.5 mL PCR tube        | Biologix, China                     |
| 1.5 mL centrifuge tube | Biologix, China                     |
| Kim wipes              | Kimberly-Clarck Worldwide Inc., USA |
| Pipette tips           | Eppendorf, Germany                  |

### 3.2 Instruments and apparatus

The list of instruments and apparatus used in this study are shown in Table 3.3.

Table 3.3: List of instruments and apparatus used in this study.

| <b>Instrument</b>                           | <b>Brand</b>               | <b>Model</b>    |
|---------------------------------------------|----------------------------|-----------------|
| Analytical Balance                          | Sartorius, Germany         | BSA 224S-CW     |
| DNA Electrophoresis System                  | Owl, USA                   | BI              |
| Hot plate and Magnetic Stirrer              | ERLA, UK                   | EMS-HP-700      |
| Laminar Air Flow Cabinet                    | ERLA, UK                   | CFM-4           |
| Microcentrifuge                             | Hettich Universal, Germany | Micro 22R       |
| Microwave Oven                              | Elba, Italy                | EMO-1706        |
| PCR Thermal Cycler                          | Applied Biosystem, USA     | Veriti 96 Well  |
| pH meter                                    | Hanna Instruments, USA     | PH211           |
| Pipette 10 $\mu$ L, 20 $\mu$ L, 100 $\mu$ L | Eppendorf, Germany         | Research plus   |
| Power Supply                                | Bio-Rad, USA               | Power-Pac 3000V |
| Shaking Waterbath                           | Memmert, Germany           | WNB14L4         |
| Spectrophotometer                           | DeNovix, USA               | DS-11           |
| UV Transilluminator                         | Uvitec, Cambridge, UK      | BTS-26.MS       |
| Vortex Mixer                                | ERLA, UK                   | EVM-6000        |

### **3.3 Chemical and reagent preparation**

#### **3.3.1 Preparation of 10X tris-borate-EDTA TBE buffer (pH 8.3)**

A total of 108 g of Tris base, 55 g boric acid and 9.3 g of disodium ethylenediaminetetraacetic acid ( $\text{Na}_2\text{EDTA}$ ) were weighted and dissolved in 800 mL of distilled water using a magnetic stirrer on a hot plate until the solution was fully mixed. After fully mixed, the solution was transferred into a 1 L Schott bottle. Distilled water was added to the solution to make the final volume of 1000 mL. The solution was then autoclaved and stored at room temperature.

#### **3.3.2 Preparation of 1X tris-borate-EDTA TBE buffer (pH 8.3)**

A total of 100 mL of 10X TBE buffer was mixed into 900 mL of distilled water in 1 L Schott bottle.

### **3.4 Methodology**

#### **3.4.1 Sample collection**

Blood samples from 100 unrelated Murut individuals were collected after interviewing them and obtaining their written consent. The sampling locations for Murut individuals were Nabawan, Sabah. A total of 10 mL of blood sample was collected from each volunteer and kept inside an EDTA containing vacutainer blood collection tube after obtaining their consent. The blood samples were stored under  $-80\text{ }^{\circ}\text{C}$  for long-term preservation. Ethical permit for this study was approved and obtained from the Human Research Ethics Committee, Universiti Sains Malaysia (USM/JEPeM/15100366).

### **3.4.2 Sampling method**

The Murut and Melanau individuals were selected through verbal interviews before being accepted as subjects for this research. All consensual participants were required to fill out the consent forms.

The inclusion and exclusion criteria of the biological sample are as follows:

Inclusion criteria:

- i. Participants must have self-declared ancestry with at least three (3) generations of similar ethnicity.
- ii. Participants should have adequate knowledge of their family lineage and background history.
- iii. Participants must be adults aged 18 years or above.

Exclusion criteria:

- i. Individuals with a history of mixed parentage or uncertainty about family lineage.
- ii. Individuals who did not provide written consent for blood collection.
- iii. Individuals with a history of diseases.