A STUDY OF MT-DNA HYPERVERSITABLE REGION 1 (HVI) SEQUENCE POLYMORPHISM ON RESIDENTS OF JOHOR IN MALAYSIA

ALIREZA LARI

UNIVERSITI TEKNOLOGI MALAYSIA
A STUDY OF MT-DNA HYPERVARIABLE REGION 1 (HVI) SEQUENCE
POLYMORPHISM ON RESIDENTS OF JOHOR IN MALAYSIA

ALIREZA LARI

A dissertation submitted in partial fulfillment of the
requirements for the award of the degree of
Master of Science (Biotechnology)

Faculty of Biosciences and Medical Engineering
Universiti Teknologi Malaysia

JANUARY 2013
This dissertation is dedicated to my family for their endless support and encouragement.
ACKNOWLEDGEMENT

I would like to express my gratitude to all those who gave me a helping hand to complete this thesis. Special thanks goes to my supervisor, Dr. Mohammed Abu Naser who is most responsible for helping me complete the thesis as well as challenging research it.

Gratitude also dedicated to all my friends. Thanks for all the helps and supports during all the way upon completing thesis. Last but not least, millions of gratitude for my beloved family, especially my father Dr. Jalil Lari and my mother Sharare Teflisi. Thanks for their love and support.
ABSTRACT

In the past few decades, lots of studies were carried out to find a way for searching relationships among various races of human beings in understanding their relatedness within a population. This study has been performed to generate DNA sequence data with a view to identifying genetic marker and using them in detecting a population in a particular area, such as, in our case, Johor, Malaysia. Mitochondrial DNA (MtDNA) sequence is usually used to detect relatedness between the races within a human population of particular region. The sequence of the region including hypervariable region 1 (HV1, nt16024-16365) and hypervariable region 2 (HV2, nt73-340) are typically used in forensic study. In this study, polymorphisms of (HV1) were investigated to discover a potential forensic marker for detecting different races. In this research PCR amplification after DNA extraction was done to detect the sequence of (HV1). Then, comparisons between Anderson reference - The Cambridge Reference Sequence (rCRS) for human mitochondrial DNA - and the sequences obtained in the current study was done. The results showed the 48 nucleotide variations within the sequence of around 340 bp nucleotides can be used as a genetic marker in forensic study among Malay population of Johor.
Dalam beberapa dekad yang lalu, banyak kajian telah dijalankan untuk mencari jalan untuk mencari hubungan di kalangan pelbagai kaum manusia dalam memahami hubungan mereka di kalangan penduduk. Kajian ini telah dilakukan untuk menjana data jujukan DNA dengan tujuan untuk mengenal pasti penanda genetik dan menggunakan mereka dalam mengesan penduduk di kawasan tertentu, seperti dalam kes kami, Johor, Malaysia. DNA Mitokondria (mtDNA) urutan biasanya digunakan untuk mengesan hubungan antara kaum di kalangan penduduk manusia di rantau tertentu. Rentetan rantau ini, termasuk rantau hypervariable 1 (HV1, nt16024-16365) dan rantau hypervariable 2 (HV2, nt73-340) biasanya digunakan dalam kajian forensik. Dalam kajian ini, polimorfisme (HV1) telah disiasat untuk menemui potensi penanda forensik untuk mengesan pelbagai kaum.

Dalam kes ini, amplifikasi PCR selepas pengekstrakan DNA telah dilakukan untuk mengesan urutan (HV1). Kemudian, perbandingan antara Anderson rujukan - Rujukan Cambridge Sequence (rCRS) bagi DNA mitokondria manusia - dan urutan yang diperolehi dalam kajian semasa telah dilakukan. Keputusan menunjukkan 48 variasi nukleotida dalam urutan sekitar 340 bp nukleotida boleh digunakan sebagai penanda genetik dalam kajian forensik di kalangan penduduk Melayu Johor.
# TABLE OF CONTENT

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>DECLARATION</td>
<td></td>
<td>iii</td>
</tr>
<tr>
<td>DEDICATION</td>
<td></td>
<td>iv</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td></td>
<td>vi</td>
</tr>
<tr>
<td>ABSTRAK</td>
<td></td>
<td>vii</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td></td>
<td>viii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td></td>
<td>xi</td>
</tr>
<tr>
<td>LIST OF ABBREVIATION</td>
<td></td>
<td>xii</td>
</tr>
</tbody>
</table>
2.5.2 Genotype of Cytocrom c 14
2.5.3 Genotype of Cytocrom b 15
2.6 Genomic Identification 15
2.7 Markers Detector 16
2.8 Mitochondrial genetic disease 17

3 RESEARCH METHODOLOGY
3.1 Experimental design 19
3.2 Selection of Sample 20
3.3 Genom DNA Extraction 20
3.4 Analysis of Genomic DNA by Gel Electrophoresis 21
3.5 Primer Designing 21
3.6 Amplification of Hypervariable Region 1 (HVI) 22
3.7 Purification of PCR Product from Gel Electrophoresis 23
3.8 Sequencing of DNA 24
3.9 Bioinformatics Analysis 24

4 RESULTS AND DISCUSSION
4.1 Genomic DNA Extraction 26
4.2 Set of Primers 28
4.3 Amplification of HVI Region by PCR 29
4.4 Purification of PCR Product 30
4.5 Sequence of HVI Region 30
4.6 Polymorphism Identification 31
4.7 Forensic Case of Study 34

5 CONCLUSIONS AND FUTURE WORK
5.1 Conclusions 36
5.2 Future Work 37

REFERENCES 38
APPENDIX 46
## LIST OF TABLE

<table>
<thead>
<tr>
<th>TABLE NO.</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Primer Sequence Used in this Research</td>
<td>22</td>
</tr>
<tr>
<td>3.2</td>
<td>Reagent for Amplification</td>
<td>23</td>
</tr>
<tr>
<td>3.3</td>
<td>PCR Amplification Program cycle</td>
<td>23</td>
</tr>
<tr>
<td>4.1</td>
<td>HVI Control Region Sequence Polymorphism</td>
<td>33</td>
</tr>
<tr>
<td>4.2</td>
<td>Characteristics HVI in Malay Population</td>
<td>34</td>
</tr>
</tbody>
</table>
# LIST OF FIGURE

<table>
<thead>
<tr>
<th>FIGURE NO.</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Mitochondrial DNA</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>Mt DNA and Displacement Loop (D-loop)</td>
<td>3</td>
</tr>
<tr>
<td>3.1</td>
<td>Experimental Procedures</td>
<td>18</td>
</tr>
<tr>
<td>4.1</td>
<td>DNA Electrophoresis</td>
<td>27</td>
</tr>
<tr>
<td>4.2</td>
<td>The Result of PCR</td>
<td>29</td>
</tr>
<tr>
<td>4.3</td>
<td>Homopolymer Regions in Mistaken Results</td>
<td>30</td>
</tr>
<tr>
<td>4.4</td>
<td>Site of Polymorphism</td>
<td>32</td>
</tr>
</tbody>
</table>
**LIST OF ABBREVIATION**

<table>
<thead>
<tr>
<th>SYMBOLS</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>bp</td>
<td>Base pair</td>
</tr>
<tr>
<td>Cyt b</td>
<td>Cytochrome b</td>
</tr>
<tr>
<td>COI</td>
<td>Cytochrome c- oxidase</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonuclease acid</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylen diamine tetraacetate</td>
</tr>
<tr>
<td>gr</td>
<td>Gram</td>
</tr>
<tr>
<td>HV1</td>
<td>hypervariable region 1</td>
</tr>
<tr>
<td>HV2</td>
<td>hypervariable region 2</td>
</tr>
<tr>
<td>kb</td>
<td>Kilobase</td>
</tr>
<tr>
<td>L</td>
<td>Litre</td>
</tr>
<tr>
<td>ml</td>
<td>Millilitre</td>
</tr>
<tr>
<td>ºC</td>
<td>Degree centigrade</td>
</tr>
<tr>
<td>%</td>
<td>Percentage</td>
</tr>
<tr>
<td>rpm</td>
<td>Revolution per minute</td>
</tr>
<tr>
<td>STR</td>
<td>Short tandem repeats</td>
</tr>
<tr>
<td>SNP</td>
<td>single-nucleotide polymorphism</td>
</tr>
<tr>
<td>SSR</td>
<td>Simple Sequence Repeats</td>
</tr>
<tr>
<td>TAE</td>
<td>Tris-acetate-EDTA</td>
</tr>
<tr>
<td>Taq</td>
<td>Termus aquaticus</td>
</tr>
<tr>
<td>µg</td>
<td>microgram</td>
</tr>
<tr>
<td>µl</td>
<td>microlitre</td>
</tr>
<tr>
<td>v/v</td>
<td>volume by volume</td>
</tr>
<tr>
<td>UTM</td>
<td>Universiti Teknologi Malaysia</td>
</tr>
</tbody>
</table>
INTRODUCTION

1.1 Background of Study

The first effort at reconstructing the evolutionary history of human population with DNA technology was initiated by Brown (1980). He studied on mitochondrial DNA (mtDNA) and revealed the presence of specific patterns of restriction enzyme cleavage on population. In 1987, Cann showed the genetic evidence that ancestor of all existing humans lived in Africa around 200,000 years ago. Since then, lots of studies have been done on the origins of human and their expansion on the earth. (Hedges et al., 1992; Nei 1992; Templeton 1992).

The present study was carried out to find out similarities and differences among the samples which can make HVI gene as a forensic marker for Malay population. Furthermore these studies will have potential application in forensic caseworks. Mitochondrial DNA sequences of a given population could be used to track down the geographic origin of that population. It has been found that there is a correlation between a specific pattern in mitochondrial DNA (mtDNA) sequences and geographic origin (Maruyama et al., 2010). Discoveries of specific mitochondrial
DNA sequences such as, Anderson sequence, for tracking down the geographic origin of a race, have been the great achievements.

Mitochondria are a tiny organelle which is located in the cytoplasm of cell and situated between the cell nucleus and the outer membrane. There are lots of mitochondria per cell \((10^3-10^4\) copies), and for each one small circle of DNA has been evolved from their distant bacterial ancestry (Maruyama et al., 2010). Mitochondrial DNA (mtDNA) has some benefits. First, their inheritance pattern is unique. The eggs of human contains a large number of mitochondria, while the sperm cells relatively lower number of mitochondria that is the numbers required to produce enough energy required during swinging towards the egg. Thus, fertilization accrues without any cooperation of male’s mitochondria thus the child except sex chromosomes which get from both parents, obtain mitochondrial DNA from their mother. Therefore, mitochondrial DNA can connect the maternal lines of living people who are living in different parts of the word.

The other characteristics of mitochondrial DNA is its ability to change more rapid (20 times faster) than nuclear DNA, due to the lack of an efficient proof-reading system. Hence cannot check the errors while DNA is copied (Sykes 2003).

Figure 1.1 Mitochondrial DNA
There is a scope for geneticists and historians to use mitochondria DNA for tracing maternal genealogies and help them to trace much deeper evolutionary connection of human race (Sykes 2003). The exciting thing about mtDNA is that, possible to discover the traces of some evolutionary relics in the cell. mtDNA replication can play the main role to protect copy number of DNA sequences. Synthesis of the heavy strand – Circular molecules of DNA, such as mitochondrial genomes which includes of two strand of DNA – initiates replication of mtDNA and lead to relocate the parental heavy strand. At 700bp downstream synthesis will be terminated, produce7S DNA (displacement loop or D-loop strand). This unexpected termination creates a characteristic triple stranded arrangement, named the D-loop (Figure 1.2) (Kasamatsu et al., 1971). Although the main function of mitochondrial D-loops is not discovered, D-loop supposed to contribute in regulation of mtDNA replication (Takamatsu et al., 2002). There is a control region in mtDNA which consist of origin of heavy strand (H strand) replication, the displacement loop (D-loop) and both origins of transcription, is the region which is too variable in the human mtDNA genome (Nei, 1995).

![mtDNA and displacement loop (D-loop)](image)
Analysis of mtDNA D-loop region has confirmed to be a best tool to identify forensic human, particularly for highly degraded samples (Cordner et al., 2011). As mentioned above, the sequence of the region including hypervariable region 1 (HV1, nt16024-16365) and hypervariable region 2 (HV2, nt73-340) are typically used for forensic studies. Furthermore, the entire D-loop sequence contains the hypervariable region 3 (HV3, nt438-576) and has been explained more information for haplotype determination - combination of DNA sequences at adjacent locations on the chromosome - (Haslindawaty et al., 2010). PCR amplification products of small mtDNA fragments with abundant polymorphisms are preferred for forensic analysis when dealing with highly degraded DNA samples due to observing different distributions of polymorphisms in mtDNA among different population groups (Budowle et al., 2003). Therefore, the important issue is evaluation of the effectiveness of each region in D-loop sequence analysis. Diverse populations groups may show vary distribution of polymorphisms in mtDNA. The important matter is large database availability with distribution in the particular population among evidence sample is essential to estimate mtDNA profile and function of mtDNA markers for the case of study on forensic (Hwa et al., 2012).

Based on technological progress in the field of Human Genome and International HapMap Projects, Single nucleotide polymorphism (SNP) is situated for short tandem repeat (STR) - DNA regions with short repeat units with usually 2-6 bp in length - and specially, for high potential of forensic DNA typing community (Butler et al., 2007). In the human genome, the most frequent form of DNA sequence variations are named as single nucleotide polymorphisms (SNPs), and are used for mapping studies, medical diagnostics, and identity testing. SNPs are Biallelic (pertaining to both alleles of a gene), which have two possible nucleotides (alleles) and only have 1% frequencies at a particular segment of human genome. However, SNPs are less polymorphic than other microsatellite markers (such as SSRs and STRs). SNPs are more abundant –approximately occur every kilobase (Cooper et al., 1985).
Usual technologies and scoring single nucleotide polymorphism are time consuming and expensive. Recently, for identifying candidate SNP segment denaturing high-performance liquid chromatography was used by Lian at 2005.

DNA bar-coding simply defined as screening of one or a few reference gene for: firstly, referring unknown individuals to species, and secondly, improving detection method for identifying new species.

1.2 Problem statement

There are many races in Malaysia. The majority of them are Malay, Chinese and Indian. The main race, Malay represents approximately 50.4% of total Malaysian population and is the focus of this current study. Finding any specific part of country for specific races is impossible due to expansion of new races among indigenous population. It is expected that each of the races has their own unique polymorphism inside the mtDNA genome. However, for distinguishing and identifying one race from the other can be challenging.

Currently, there is not much research done on SNP markers for Malay population in the literature. It has been assumed in this current study that there might be exist a specific mitochondrial DNA sequence which can potentially be used for identifying Malay races. An attempt has been made to explore whether the hypervariable region I (HVI) of mtDNA can be used as a marker for detecting variation of different races among Malaysia population.
1.3 Objectives of Study

I. To amplify and sequence the HV1 gene using PCR and sequencer, respectively.

II. To compare obtained sequence data within the samples as well as with the publicly available sequence data for reference sequence.

III. To analyze the polymorphism in HVI gene in a given population and to investigate the usefulness of this polymorphic HVI in Malay race identification.

1.4 Scope of Study

The scope of this study consists of three parts. The first is the collection of blood samples from fifteen Malay individual living in Johor. The blood samples with permission were collected from Universiti Teknologi Malaysia (UTM) Health Center, and background information of each individual was also recorded. The second part was the laboratory works, such as DNA extraction from blood sample; PCR amplification using designed primer for HVI gene; and sequencing. The third part was the analyzing of sequence data using bioinformatics tools such as sequence alignments of hypervariable region I (HVI) with publicly available database, development of DNA barcode for Malay population by Genius software.
1.5 Significance of Study

In this study, method for identifying races using mtDNA HVI genome were initiated to ease study of biodiversity and forensic analysis. Since Cytochrom $c$-oxidase and $b$ oxidase gene from mitochondria not appropriate due to high length of sequences and not much variable region, hypervariable region 1 (HVI) expected as potentially use for genomic analysis. The polymorphisms study helpful in forensic studies.
REFERENCES


