GENETIC COMPARISON OF DIFFERENT HUMAN POPULATION GROUPS
USING INTERNAL TRANSCRIBED SPACER SEQUENCE

MUNA I. HADI G. ALATIYAH

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
To my beloved mother and father
ACKNOWLEDGEMENT

First, Alhamdulillah, thanks to Allah for finishing the research and I would like to express my sincere thanks and appreciation to my supervisor Assoc.Prof Dr. Mohd. Shahir Shamsir Omar, who generously guided me with his knowledge and experience in every step of my thesis. My special thanks to Dr. Salehuddin Bin Hamdan and I would like to take this opportunity to deliver my appreciation to him for letting me work in animal tissue culture laboratory and special thanks to all technical staff members in the bioinformatics and animal tissue culture laboratories especially Siti Nur Hafizah binti Soid at UTM for their assistance and collaboration during my laboratory work. Thanks to all my friends who showed their concern and support to me. Their views and advices are useful indeed.
ABSTRACT

Genetic identity testing for humans has been used to examine the variations in the polymorphic regions of the human DNA. These methods include the RFLP (Restriction fragment length polymorphisms), the STR (short tandem repeats) markers that are most commonly used loci for human identification as well as the mtDNA or Y chromosome in forensic medicine and paternity tests. Currently, the internal transcribed spacer (ITS) region is used for the evolutionary analysis of different species of animals, plants, fungi, yeast. However, the ITS region is yet to be used in genetic identity testing for humans. In this study, the genetic comparison of different human population groups using ITS sequence of mtDNA was performed. Two segments of the region of human mtDNA were selected and sequenced in order to determine if any sufficient single nucleotide polymorphisms (SNPs) exist and if it is suitable for genetic identity testing in human. Specific primers were designed to amplify the ITS regions using PCR from sample extracted from the blood samples of different nationalities of students from Faculty of Biosciences and Bioengineering, UTM (Iran, Nigeria, Kurdistan, Malaysia, Palestine and Luxembourg). The result from the bioinformatics analysis showed a significant number of SNPs in the ITS region. There are 40 SNPs found in the ITS sequences and 35 SNPs in the NADH1 sequences of the mtDNA. The phylogenetic analysis of the result revealed the phylogenetic relationship and a distinct genetic difference between the different nationalities that may be used as a marker for human genetic identification in the future.
ABSTRAK

Ujian identiti genetik untuk manusia telah digunakan untuk mengkaji variasi dalam kawasan polimorfik DNA manusia. Kaedah-kaedah ini termasuk penanda RFLP (Sekatan serpihan panjang polimorfisme), penanda lokus STR (pengulangan selaras pendek) yang paling biasa digunakan dalam pengenalpastian lokus manusia serta kromosom mtDNA atau Y dalam perubatan forensik dan ujian paterniti. Sekarang, rantau salinan peruang dalaman (ITS) digunakan untuk menjalankan analisis evolusi bagi pelbagai spesies haiwan, tumbuhan, kulat, yis. Walau bagaimanapun, rantau ITS belum lagi digunakan dalam ujian identiti genetik manusia. Dalam kajian ini, perbandingan genetik bagi populasi manusia yang berbeza dengan menggunakan jujukan ITS mtDNA telah dilakukan. Dua segmen rantau mtDNA manusia telah dipilih dan dijujukan dalam usaha untuk menentukan kewujudan polimorfisme nukleotida tunggal (SNP) yang mencukupi dan kesesuaianannya untuk ujian identiti genetik dalam manusia. Primers khusus telah direka untuk mengamplifikasikan rantau ITS dengan menggunakan PCR untuk mengekstrak sampel darah pelajar dari Fakulti Biosains dan Biokejuruteraan, UTM yang terdiri daripada kewarganegaraan yang berbeza (Iran, Nigeria, Kurdistan, Malaysia, Palestin dan Luxembourg). Hasil daripada analisis bioinformatik menunjukkan terdapat jumlah tererti SNPs di rantau ITS. Terdapat 40 SNPs ditemui di jujukan ITS dan 35 SNPs dalam jujukan NADH1 mtDNA. Keputusan analisis filogenetik memperlihatkan hubungan filogeni dan perbezaan genetik yang ketara antara bangsa yang berbeza yang boleh digunakan sebagai penanda bagi pengenalan genetik manusia pada masa depan.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>DECLARATION</td>
<td>iii</td>
<td></td>
</tr>
<tr>
<td>DEDICATION</td>
<td>iv</td>
<td></td>
</tr>
<tr>
<td>ACKNOWLEDGEMENT</td>
<td>vi</td>
<td></td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>vii</td>
<td></td>
</tr>
<tr>
<td>ABSTRAK</td>
<td>vii</td>
<td></td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>ix</td>
<td></td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xi</td>
<td></td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xii</td>
<td></td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>xiv</td>
<td></td>
</tr>
<tr>
<td>LIST OF APPENDICES</td>
<td>xvi</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1</td>
<td>Background of Study</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>Problem Statement</td>
<td>2</td>
</tr>
<tr>
<td>1.3</td>
<td>The aim of the study</td>
<td>2</td>
</tr>
<tr>
<td>1.4</td>
<td>Objectives</td>
<td>3</td>
</tr>
<tr>
<td>1.5</td>
<td>Significant of Study</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>LITERATURE REVIEW</td>
<td>5</td>
</tr>
<tr>
<td>2.1</td>
<td>Genetic identification markers</td>
<td>5</td>
</tr>
<tr>
<td>2.2</td>
<td>Single nucleotide polymorphisms</td>
<td>8</td>
</tr>
<tr>
<td>2.3</td>
<td>mtDNA diversity in humans</td>
<td>9</td>
</tr>
<tr>
<td>2.4</td>
<td>Ribosomal DNA</td>
<td>12</td>
</tr>
</tbody>
</table>
3 RESEARCH METHODOLOGY 15
3.1 The flow chart of research 15
3.2 Bioinformatics lab work 16
   3.2.1 Identification of sequence of ITS region 16
   3.2.2 Designing the primers for sequence amplification 17
3.3 Blood sample collection 18
3.4 Extraction of Genomic DNA from whole blood 18
3.5 Polymerase Chain Reaction (PCR) 19
3.6 Qualitative analysis of PCR products 21
3.7 Purification of PCR products 24
3.8 DNA sequencing 25
3.9 Phylogenetic analysis 25

4 RESULT AND DISCUSSION 26
4.1 Bioinformatics work 26
   4.1.1 Identification of the conserved sequence of the ITS region 26
   4.1.2 Primer design 32
4.2 Extraction of Genomic DNA from whole blood 36
4.3 Sequence amplification using Polymerase Chain Reaction (PCR) 37
4.4 The purification of PCR products 39
4.5 Qualitative analysis of purified PCR products 40
4.6 Phylogenetic analysis 42
4.7 Phylogenetic tree 51

5 CONCLUSION AND FUTURE WORK 54
5.1 Conclusion 54
5.2 Future work 55

REFERENCES 56

APPENDICES 61
LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE NO.</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>The human mtDNA sequences from (NCBI) database</td>
<td>16</td>
</tr>
<tr>
<td>3.2</td>
<td>PCR master mix reaction of 50 μl to amplifying DNA template</td>
<td>20</td>
</tr>
<tr>
<td>3.3</td>
<td>Gradient temperature of PCR to detect annealing temperature</td>
<td>21</td>
</tr>
<tr>
<td>3.4</td>
<td>Cycling condition for PCR reaction</td>
<td>22</td>
</tr>
<tr>
<td>3.5</td>
<td>PCR master mix reaction of 50 μl to amplifying DNA template</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>ITS1 region</td>
<td></td>
</tr>
<tr>
<td>3.6</td>
<td>PCR reaction mixture for 50 μl to amplifying DNA template</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>NADH1 region</td>
<td></td>
</tr>
<tr>
<td>4.1</td>
<td>The primer parameter values as determined by the Oligo Calculator Software</td>
<td>35</td>
</tr>
<tr>
<td>4.2</td>
<td>The best primers designed using the Noe primerDemo software</td>
<td>35</td>
</tr>
<tr>
<td>4.3</td>
<td>Genomic DNA extraction results by using NanoDrop spectrophotometer</td>
<td>36</td>
</tr>
<tr>
<td>4.4</td>
<td>The data of Primers (For &amp; Rev) from 1st Base Company</td>
<td>37</td>
</tr>
<tr>
<td>4.5</td>
<td>The results of purified PCR products of ITS1 gene</td>
<td>39</td>
</tr>
<tr>
<td>4.6</td>
<td>The results of purified PCR products of NADH1 gene</td>
<td>40</td>
</tr>
<tr>
<td>4.7</td>
<td>The percentage identity results of ITS1 by using SIAS software</td>
<td>47</td>
</tr>
<tr>
<td>4.8</td>
<td>The percentage identity results of NADH1 by using SIAS software</td>
<td>48</td>
</tr>
<tr>
<td>4.9</td>
<td>The SNPs results of ITS1 region</td>
<td>49</td>
</tr>
<tr>
<td>4.10</td>
<td>The SNPs results of NADH1 region</td>
<td>50</td>
</tr>
<tr>
<td>FIGURE NO.</td>
<td>TITLE</td>
<td>PAGE</td>
</tr>
<tr>
<td>------------</td>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>2.1</td>
<td>RFLP (Restriction fragment length polymorphisms)</td>
<td>7</td>
</tr>
<tr>
<td>2.2</td>
<td>Mitochondrial DNA of human</td>
<td>11</td>
</tr>
<tr>
<td>2.3</td>
<td>Human nuclear ribosomal (rDNA) genes</td>
<td>13</td>
</tr>
<tr>
<td>3.1</td>
<td>Overall flow chart of research</td>
<td>15</td>
</tr>
<tr>
<td>3.2</td>
<td>Noe primerDemo Software for primers design</td>
<td>17</td>
</tr>
<tr>
<td>4.1</td>
<td>The Multiple Sequence Alignment of the ITS1 of mtDNA</td>
<td>27-30</td>
</tr>
<tr>
<td>4.2</td>
<td>The alignment of nuclear ITS2 sequence with mtDNA sequence from NCBI</td>
<td>31-32</td>
</tr>
<tr>
<td>4.3</td>
<td>A screenshot of the Noeprimer Demo software of forward and reverse primers generation of the ITS1 sequence</td>
<td>33</td>
</tr>
<tr>
<td>4.4</td>
<td>A screenshot of the Noeprimer Demo software of forward and reverse primers generation of the NADH1 sequence</td>
<td>34</td>
</tr>
<tr>
<td>4.5</td>
<td>The results of annealing temperature test in gradient PCR detected by Agarose gel electrophoresis</td>
<td>38</td>
</tr>
<tr>
<td>4.6</td>
<td>Qualitative analysis of purified PCR products of ITS1 region by using gel electrophoresis</td>
<td>41</td>
</tr>
<tr>
<td>4.7</td>
<td>Qualitative analysis of purified PCR products of NADH1 region by using gel electrophoresis</td>
<td>42</td>
</tr>
<tr>
<td>4.8</td>
<td>The results of multiple sequence alignment with Jalview software for ITS1 sequences</td>
<td>43-44</td>
</tr>
<tr>
<td>4.9</td>
<td>The results of multiple sequence alignment with Jalview software for NADH1 sequences</td>
<td>45-46</td>
</tr>
<tr>
<td>Section</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>4.10</td>
<td>Neighbour joining tree of ITS1 sequences of different populations</td>
<td>52</td>
</tr>
<tr>
<td>4.11</td>
<td>Average distance tree of ITS1 sequences of different populations</td>
<td>52</td>
</tr>
<tr>
<td>4.12</td>
<td>Neighbour joining trees of NADH1 sequences of different populations</td>
<td>53</td>
</tr>
<tr>
<td>4.13</td>
<td>Average distance tree of NADH1 sequences of different population.</td>
<td>53</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>µl</td>
<td>Microliter</td>
<td></td>
</tr>
<tr>
<td>µM</td>
<td>Micromole</td>
<td></td>
</tr>
<tr>
<td>µg</td>
<td>Microgram</td>
<td></td>
</tr>
<tr>
<td>ng</td>
<td>Nanogram</td>
<td></td>
</tr>
<tr>
<td>bp</td>
<td>Base pair</td>
<td></td>
</tr>
<tr>
<td>°C</td>
<td>Degree Centigrade Celsius</td>
<td></td>
</tr>
<tr>
<td>conc.</td>
<td>Concentration</td>
<td></td>
</tr>
<tr>
<td>D.W</td>
<td>Distilled water</td>
<td></td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
<td></td>
</tr>
<tr>
<td>dNTPs</td>
<td>Deoxnucleotide triphosphate</td>
<td></td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic Acid</td>
<td></td>
</tr>
<tr>
<td>Fbb</td>
<td>Faculty of Biosciences and Bioengineering</td>
<td></td>
</tr>
<tr>
<td>For</td>
<td>Forward</td>
<td></td>
</tr>
<tr>
<td>g</td>
<td>Gram</td>
<td></td>
</tr>
<tr>
<td>hr</td>
<td>hour</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>Identification Number</td>
<td></td>
</tr>
<tr>
<td>ITS1</td>
<td>Internal TRANSCRIBED spacer</td>
<td></td>
</tr>
<tr>
<td>ml</td>
<td>Milliliter</td>
<td></td>
</tr>
<tr>
<td>MW</td>
<td>Molecular Weight</td>
<td></td>
</tr>
<tr>
<td>mtDNA</td>
<td>Mitochondrial DNA</td>
<td></td>
</tr>
<tr>
<td>min</td>
<td>Minute</td>
<td></td>
</tr>
<tr>
<td>mg</td>
<td>Milligram</td>
<td></td>
</tr>
<tr>
<td>N.A.</td>
<td>Nothing at all</td>
<td></td>
</tr>
<tr>
<td>Nb</td>
<td>Number</td>
<td></td>
</tr>
<tr>
<td>nm</td>
<td>Monometer</td>
<td></td>
</tr>
</tbody>
</table>
NCBI - National Center for Biotechnology Information
N.C. - Negative Control
No. - Number
OD - Optical Density
OXPHOS - Oxidative Phosphorylation
PCR - Polymerase chain reaction
PID - Percentage Identity between the two sequences at each aligned position.
RNase - ribonucleases
Rev - Reveres
rpm - Revolution Per minute
Sec - Second
TAE - Tris-acetate EDTA
Tm - Melting temperature
tRNA - transfer RNA
UTM - Universiti Technology Malaysia
UV - Ultra Violet
# LIST OF APPENDICES

<table>
<thead>
<tr>
<th>APPENDIX NO.</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Genomic DNA extraction results by using NanoDrop spectrophotometer</td>
<td>61</td>
</tr>
<tr>
<td>B</td>
<td>Oligo calculator software results</td>
<td>66</td>
</tr>
<tr>
<td>C</td>
<td>Multiple sequence alignment results with Jalview software</td>
<td>71</td>
</tr>
<tr>
<td>D</td>
<td>Phylogenetic tree results using PID from Clustal W alignment of Jalview software</td>
<td>77</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

1.1 Background of Study

Genetic identity testing for humans has been used to examine the variations in the polymorphic regions of the human DNA. These methods use unique genetic markers such as the RFLP (Restriction fragment length polymorphisms), the STR (short tandem repeats), the mtDNA or Y chromosome for human identification, forensic medicine and paternity testing (Goodwin, Linacre, & Hadi, 2011).

The ribosomal DNA (rDNA) contains unique tandem repeats has been proposed as a new DNA fingerprinting source for human genetic identification. There are five chromosomes in the human genome that contain the rDNA chromosome: 13, 14, 15, 21 and 22. The ribosomal DNA consists of NTS (non-transcript spacer), 18S, ITS1 (internal transcribed spacer), 5.8S, ITS2 and 28S. The coding region with the gene is in 18S, 5.8S and 28S. The ITS (internal transcribed spacer) found between 18S, 5.8S regions and between 5.8S, 28S regions (Nei & Rooney, 2005).
Currently, the internal transcribed spacer (ITS) region has been used for studying the phylogenetic analyses of the evolution of tandemly arranged genes in different species (animals, plants, fungi and yeast) but has not yet been used in genetic identity testing in humans (Abrahams & Benjeddou, 2009).

1.2 Problem statement

Currently, the ITS region has not yet being used in genetic identity test in human. The tandemly repeated genes of ITS in rDNA are suitable candidate loci for molecular evolutionary marker studies because of their universal presence, high copy number and functional similarity. The presence of short repeated motif, high GC content and a higher change rate than other gene variable region make this an attractive option compared to existing identification markers (Blouin, 2002).

1.3 The aim of the study

This study investigates whether the variations of the ITS region of mtDNA can be used for human genetic identity testing, forensic medicine or paternity analysis. Different samples sourced from UTM FBB students were for this purpose. Students from different ethnic groups were used as examples to test this hypothesis. The sequencing and analysis of ITS of rDNA were used to construct the phylogenetic tree of the different UTM FBB
students to determine the relationship between groups of different populations. The aim of this study was to determine if the variation or polymorphic loci in ITS is sufficient to differentiate between different population groups and to detect possible sequence motifs associated with different ethnic groups.

1.4 Objectives

- To detect all the genome variations and the conserved regions of the mtDNA from existing genomic databases.
- To isolate and purify the maximum mtDNA from the whole blood sample from representations from different ethnic groups (Malay, Chinese, Nigerian, Palestine, Iranian and Luxemburg).
- To determine the most efficient and the best primers for successful amplification of region of interest in human mtDNA.
- To analyze the genetic variations using multiple sequence alignment and phylogenetic tree between the different samples of population group.

1.5 Significance of Study

The ITS region with the polymorphic loci or the variation in mtDNA in human can become a useful marker as the variability of the ITS region of rDNA that are due to
mutation, deletion or insertion in mtDNA (Gonzalez, Sylvester, Smith, Stambolian, & Schmickel, 1990) should be sufficient for the use as genetic identification. The ITS region has not been used before for human genetic identity and success in this project would create a potential novel application in human genetic identification and paternity analysis.
REFERENCES


