STATISTICAL EVIDENCE TO SUPPORT THE UTILIZATION OF 21 AUTOSOMAL SHORT TANDEM REPEATS LOCI FOR HUMAN IDENTIFICATION IN MALAYSIAN POPULATIONS

MOHD NOR AZLAN BIN AB RASHID

A dissertation submitted in partial fulfilment of the requirements for the award of the degree of Master of Science

Faculty of Science
Universiti Teknologi Malaysia

MARCH 2019
DEDICATION

This dissertation is dedicated to my father, who taught me that the best kind of knowledge to have is that which is learned for its own sake. It is also dedicated to my mother, who taught me that even the largest task can be accomplished if it is done one step at a time.
ACKNOWLEDGEMENT

In the name of Allah, the Most Beneficent, the Most Merciful. Firstly, I would like to express gratitude to Allah for his blessing until completing this research. My deepest appreciation goes to my supervisor, Dr. Naji Arafat bin Mahat for all his valuable guidance, availability, support and encouragement during my study and research. He helped me to understand my research better. Without his supervision and persistent help, this research would not have been possible.

I would like to thank SAC Dato’ Hussein bin Omar Khan, Principal Assistant Director of DNA Databank Division (D13) of Royal Malaysia Police for his permission and providing me with all the necessary data for my research. He was willing to help despite his busy schedule. I also wish to express my sincere thanks to the Vice Chancellor, Dean of School of Graduate Studies, Dean of Faculty of Science, Head of Chemistry Department, coordinator of the program and all the staff of Universiti Teknologi Malaysia for all their support and assistance during my academic session. Not forgetting to The Public Service Department of Malaysia for allowing me to further my master’s degree with paid leave. Thank you for all the opportunities and for making valuable experiences more than I ever expected.

In addition, I must also thank Puan Syarifah Nany Rahayu Kamilia and Puan Aedrianee Reeza for teaching me and clearly a doubt about my research. Special thanks go to my several friends and colleagues who helped and gave me their full commitment in completing this study. Finally, I would like to thank my beloved wife, daughters, parents and siblings for their encouragement and always motivated me during the study. Over the duration of this study, they have been sacrificed a lot since I started studying here and I am really appreciating it.
ABSTRACT

The use of 21 autosomal STRs loci for human identification has been gaining popularity throughout the world, attributable to its presumably better statistical discrimination ability. It has been indicated that the forensic statistical parameters for supporting the use of 21 STRs loci varied among the different populations, and such data for the diverse Malaysian populations remain unreported, rendering doubts in the court of law about its real ability for human identification for Malaysian population. Using the GlobalFiler™ Express PCR Amplification kit, the complete DNA profiles of 21 STRs loci from buccal swabs of convicted Malaysian criminal (n = 570; 190 each for Malays, Chinese and Indians) (2016 – 2017) were analysed for their allele frequencies, exact test of Hardy-Weinberg equilibrium, observed and expected heterozygosity, power of discrimination, power of exclusion, match probability and polymorphism information content. Being the most informative locus, SE33 demonstrated the highest power of discrimination and power of exclusion, indicating it usefulness to discriminate individuals. In contrast, TPOX had the lowest power of discrimination and power of exclusion, as well as being the less informative genetic locus for all Malaysian population studied here. The probabilities that two individuals would share the same DNA profiles among the Malaysian Malays, Chinese and Indians, as well as in general Malaysian population were $1.3713 \times 10^{-25}$, $2.8822 \times 10^{-25}$, $7.5668 \times 10^{-26}$ and $1.0385 \times 10^{-26}$, respectively. The results obtained here were found comparable with similar studies reported in other populations, hence, its robustness for forensic human identification among the Malaysian populations is therefore, statistically supported. Phylogenetic analysis was performed to determine the genetic relationship of the Malaysian population with other Asian populations.
ABSTRAK

Penggunaan 21 STRs autosom untuk pengecaman manusia semakin popular digunakan di seluruh dunia kerana mempunyai keupayaan diskriminasi yang lebih baik secara statistik. Kajian telah melaporkan mengenai parameter statistik forensik untuk menyokong penggunaan 21 STR loci di antara pelbagai populasi yang berbeza sebelum ini, namun data sedemikian untuk kepelbagaian populasi di Malaysia masih belum dilaporkan sehingga menimbulkan keraguan dari segi undang-undang di mahkamah tentang keupayaan sebenar untuk pengenalan manusia bagi populasi di Malaysia. Dengan menggunakan kit amplifikasi GlobalFiler™ Express PCR, profil DNA lengkap bagi 21 STRs autosom daripada sampel bukal pesalah yang disabitkan kesalahan di Malaysia (n = 570; 190 setiap orang untuk orang Melayu, Cina dan India) (2016 - 2017) digunakan untuk menganalisis frekuensi alel, ujian keseimbangan Hardy-Weinberg, heterozigositi yang diperhatikan dan dijangka, kuasa diskriminasi, kuasa pengecualian, kebarangkalian kepadanan dan kandungan informasi polimorfisme. Lokus SE33 merupakan lokus yang paling bermaklumat kerana menunjukkan kekuatan diskriminasi dan kuasa pengecualian yang tertinggi membolehkannya digunakan untuk mendiskriminasikan individu. Sebaliknya, TPOX pula mempunyai nilai kuasa diskriminasi dan pengecualian yang paling rendah, dan juga lokus genetik yang paling kurang bermaklumat untuk semua populasi yang dikaji di Malaysia. Kebarangkalian bahawa dua individu akan berkongsi profil DNA yang sama dalam kalangan populasi Melayu, Cina dan India di Malaysia, dan juga penduduk Malaysia pada umumnya masing-masing adalah 1.3713 x 10^{-25}, 2.8822 x 10^{-25}, 7.5668 x 10^{-26} dan 1.0385 x 10^{-26}. Hasil kajian menunjukkan bahawa kekuatan dan keupayaan untuk pengecaman manusia dalam kalangan penduduk di Malaysia adalah setanding dengan kajian yang serupa ke atas populasi di negara lain yang pernah dilaporkan dan disokong sepenuhnya secara statistik. Analisis filogenetik juga telah dilakukan untuk menentukan hubungan genetik antara populasi Malaysia dengan populasi negara-negara Asia yang lain.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>DECLARATION</td>
<td>ii</td>
</tr>
<tr>
<td>DEDICATION</td>
<td>iii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENT</td>
<td>iv</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>v</td>
</tr>
<tr>
<td>ABSTRAK</td>
<td>vi</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xiv</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>xv</td>
</tr>
<tr>
<td>LIST OF SYMBOLS</td>
<td>xvii</td>
</tr>
<tr>
<td>LIST OF APPENDICES</td>
<td>xviii</td>
</tr>
</tbody>
</table>

## CHAPTER 1  INTRODUCTION

1.1 Research Background  
1.2 Problem Statement  
1.3 Objectives and Hypothesis  
1.4 Scope of Research  
1.5 Significance of Study

## CHAPTER 2  LITERATURE REVIEW

2.1 Deoxyribonucleic Acid (DNA)  
2.2 DNA in Forensic Science  
2.3 Molecular Markers for Human Identification  
   2.3.1 Restriction Fragment Length Polymorphisms (RFLP)  
   2.3.2 Short Tandem Repeat (STR)  
   2.3.3 Single Nucleotide Polymorphism (SNP)  
2.4 The Evolution of STRs Multiplexes
2.5 GlobalFiler™ Express PCR Amplification kit 14
2.6 Admissibility of DNA Evidence in Court 16
2.7 DNA Databases 17
2.8 Population Studies 19
  2.8.1 Malaysian Population 19
  2.8.2 Forensic Statistical Parameter 21
    2.8.2.1 Hardy-Weinberg Equilibrium 21
    2.8.2.2 Power of Discrimination, Power of Exclusion and Matching Probability 21
    2.8.2.3 Polymorphism Information Content 22

CHAPTER 3 RESEARCH METHODOLOGY 23
3.1 Materials 23
3.2 Methods 26
  3.2.1 Experimental Design 26
  3.2.2 Collection of Samples 27
  3.2.3 DNA Extraction 28
  3.2.4 PCR Amplification 28
  3.2.5 Capillary Electrophoresis 29
  3.2.6 Laboratory Waste Disposal 29
  3.2.7 Forensic Statistical Parameter Analyses 30
    3.2.7.1 Allele frequency 30
    3.2.7.2 Observed Heterozigosity (Ho), Expected Heterozigosity (He), Hardy–Weinberg equilibrium (HWE) 30
  3.2.7.3 Power of Discrimination (PD) 31
  3.2.7.4 Power of Exclusion (PE) 32
  3.2.7.5 Matching Probability (MP) 33
  3.2.7.6 Polymorphism Information Content (PIC) 33
  3.2.8 Phylogenetic Tree 34
CHAPTER 4  RESULTS AND DISCUSSIONS  35
4.1 Dynamics of Malaysian Population and Rationale  35
4.2 Allele Frequency  36
  4.2.1 Locus D3S1358  83
  4.2.2 Locus vWA  84
  4.2.3 Locus D16S539  85
  4.2.4 Locus CSF1PO  86
  4.2.5 Locus TPOX  88
  4.2.6 Locus D8S1179  89
  4.2.7 Locus D21S11  90
  4.2.8 Locus D18S51  91
  4.2.9 Locus D19S433  92
  4.2.10 Locus TH01  93
  4.2.11 Locus FGA  94
  4.2.12 Locus D5S818  96
  4.2.13 Locus D13S317  97
  4.2.14 Locus D7S820  98
  4.2.15 Locus D2S1338  99
  4.2.16 Loci D2S441, D22S1045 and D10S1248  100
  4.2.17 Loci SE33, D1S1656 and D12S391  102
4.3 DNA Statistical Parameter  106
  4.3.1 Hardy-Weinberg Equilibrium  106
  4.3.2 Heterozygosity  107
  4.3.3 Power of Discrimination and Combined Power of Discrimination  108
  4.3.4 Power of Exclusion and Combined Power of Exclusion  110
  4.3.5 Match Probability and Combined Match Probability  110
  4.3.6 Polymorphism Information Content  112
  4.3.7 Phylogenetic tree  113

CHAPTER 5  CONCLUSION AND RECOMMENDATIONS  115
5.1 Conclusion  115
<table>
<thead>
<tr>
<th>TABLE NO.</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2.1</td>
<td>Information on Commonly Used Autosomal STR Loci. The 13 CODIS Core Loci Are Highlighted in Bold Font</td>
<td>15</td>
</tr>
<tr>
<td>Table 2.2</td>
<td>Ethnic groups of Malaysian Population.</td>
<td>20</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>List of consumables and laboratory ware.</td>
<td>24</td>
</tr>
<tr>
<td>Table 3.2</td>
<td>List of chemicals and reagents</td>
<td>24</td>
</tr>
<tr>
<td>Table 3.3</td>
<td>List of laboratory instruments.</td>
<td>25</td>
</tr>
<tr>
<td>Table 3.4</td>
<td>List of Software</td>
<td>25</td>
</tr>
<tr>
<td>Table 3.5</td>
<td>Optimized PCR thermal cycling protocol.</td>
<td>28</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>Allele frequencies and the forensic statistical parameters for 21 autosomal STR loci for the Malaysian Malay population</td>
<td>37</td>
</tr>
<tr>
<td>Table 4.2</td>
<td>Allele frequencies and the forensic statistical parameters derived from 21 autosomal STRs loci for the Malaysian Chinese population</td>
<td>47</td>
</tr>
<tr>
<td>Table 4.3</td>
<td>Allele frequencies and the forensic statistical parameters derived from 21 autosomal STRs loci for the Malaysian Indian population.</td>
<td>57</td>
</tr>
<tr>
<td>Table 4.4</td>
<td>Allele frequencies and the forensic statistical parameters derived from 21 autosomal STRs loci for the combined ethnic groups (Malays, Chinese and Indians).</td>
<td>69</td>
</tr>
<tr>
<td>Table 4.5</td>
<td>Allele frequencies of STR D3S1358 genetic locus for the Malaysian Malays, Chinese and Indians, as well as their combination into a single Malaysian population.</td>
<td>84</td>
</tr>
<tr>
<td>Table 4.6</td>
<td>Allele frequencies of STR vWA genetic locus for the Malaysian Malays, Chinese and Indians, as well as their combination into a single Malaysian population.</td>
<td>85</td>
</tr>
<tr>
<td>Table 4.7</td>
<td>Allele frequencies of STR D16S539 genetic locus for the Malaysian Malays, Chinese and Indians, as well as their combination into a single Malaysian population.</td>
<td>86</td>
</tr>
<tr>
<td>Table 4.8</td>
<td>Allele frequencies of STR CSF1PO genetic locus for the Malaysian Malays, Chinese and Indians, as well as their combination into a single Malaysian population.</td>
<td>87</td>
</tr>
</tbody>
</table>
Table 4.9  Allele frequencies of STR TPOX genetic locus for the Malaysian Malays, Chinese and Indians, as well as their combination into a single Malaysian population.

Table 4.10  Allele frequencies of STR D8S1179 genetic locus for the Malaysian Malays, Chinese and Indians, as well as their combination into a single Malaysian population.

Table 4.11  Allele frequencies of STR D21S11 genetic locus for the Malaysian Malays, Chinese and Indians, as well as their combination into a single Malaysian population.

Table 4.12  Allele frequencies of STR D18S51 genetic locus for the Malaysian Malays, Chinese and Indians, as well as their combination into a single Malaysian population.

Table 4.13  Allele frequencies of STR D19S433 genetic locus for the Malaysian Malays, Chinese and Indians, as well as their combination into a single Malaysian population.

Table 4.14  Allele frequencies of STR TH01 genetic locus for the Malaysian Malays, Chinese and Indians, as well as their combination into a single Malaysian population.

Table 4.15  Allele frequencies of STR FGA genetic locus for the Malaysian Malays, Chinese and Indians, as well as their combination into a single Malaysian population.

Table 4.16  Allele frequencies of STR D5S818 genetic locus for the Malaysian Malays, Chinese and Indians, as well as their combination into a single Malaysian population.

Table 4.17  Allele frequencies of STR D13S317 genetic locus for the Malaysian Malays, Chinese and Indians, as well as their combination into a single Malaysian population.

Table 4.18  Allele frequencies of STR D7S820 genetic locus for the Malaysian Malays, Chinese and Indians, as well as their combination into a single Malaysian population.

Table 4.19  Allele frequencies of STR D2S1338 genetic locus for the Malaysian Malays, Chinese and Indians, as well as their combination into a single Malaysian population.

Table 4.20  Allele frequencies of STR D2S441 genetic locus for the Malaysian Malays, Chinese and Indians, as well as their combination into a single Malaysian population.

Table 4.21  Allele frequencies of STR D22S1045 genetic locus for the Malaysian Malays, Chinese and Indians, as well as their combination into a single Malaysian population.
Table 4.22  Allele frequencies of STR D10S1248 genetic locus for the Malaysian Malays, Chinese and Indians, as well as their combination into a single Malaysian population.

Table 4.23  Allele frequencies of STR SE33 genetic locus for the Malaysian Malays, Chinese and Indians, as well as their combination into a single Malaysian population.

Table 4.24  Allele frequencies of STR D1S1656 genetic locus for the Malaysian Malays, Chinese and Indians, as well as their combination into a single Malaysian population.

Table 4.25  Allele frequencies of STR D12S391 genetic locus for the Malaysian Malays, Chinese and Indians, as well as their combination into a single Malaysian population.

Table 4.26  The combined power of discrimination, combined power of exclusion and combined match probability of Malaysian Malay, Chinese and Indian, as well as in combination.
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE NO.</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>The conceptual framework of this present research</td>
<td>4</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>Overall flow chart of this research</td>
<td>27</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>Representative electropherogram for the DNA profile of an individual analysed in this present research.</td>
<td>81</td>
</tr>
<tr>
<td>Figure 4.2</td>
<td>Neighbour-joining tree between Malaysian Malays, Chinese, Indian, and combination (Malaysian Malay, Chinese and Indian) together with others Asian populations constructed based on genetic distances calculated using the allele frequencies of 21 autosomal STRs genetic loci. (Numbers above branches are node support after the bootstrap technique in PHYLIP package programme SEQBOOT with 1000 replicates).</td>
<td>114</td>
</tr>
</tbody>
</table>
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABI</td>
<td>Applied Biosystems</td>
</tr>
<tr>
<td>B.C</td>
<td>Before century</td>
</tr>
<tr>
<td>BDFFM</td>
<td>‘Bahagian Databank DNA Forensik Malaysia’</td>
</tr>
<tr>
<td>bp</td>
<td>Base pairs</td>
</tr>
<tr>
<td>CMP</td>
<td>Combined Match Probability</td>
</tr>
<tr>
<td>CPD</td>
<td>Combined Power of Discrimination</td>
</tr>
<tr>
<td>CPE</td>
<td>Combined Power of Exclusion</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic Acid</td>
</tr>
<tr>
<td>DNA BASES</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>Adenine</td>
</tr>
<tr>
<td>T</td>
<td>Thymine</td>
</tr>
<tr>
<td>C</td>
<td>Cytosine</td>
</tr>
<tr>
<td>G</td>
<td>Guanine</td>
</tr>
<tr>
<td>FDDM</td>
<td>Forensic DNA Databank Malaysia</td>
</tr>
<tr>
<td>FSS</td>
<td>Forensic Science Service</td>
</tr>
<tr>
<td>He</td>
<td>Expected Heterozygosity</td>
</tr>
<tr>
<td>Ho</td>
<td>Observed Heterozygosity</td>
</tr>
<tr>
<td>HWE</td>
<td>Hardy-Weinberg Equilibrium</td>
</tr>
<tr>
<td>i.e.</td>
<td>In Essence</td>
</tr>
<tr>
<td>MP</td>
<td>Match Probability</td>
</tr>
<tr>
<td>MPCHR</td>
<td>Multiplex Polymerase Chain Reaction</td>
</tr>
<tr>
<td>MSDS</td>
<td>Material Safety Data Sheet</td>
</tr>
<tr>
<td>mtDNA</td>
<td>Mitochondrial DNA</td>
</tr>
<tr>
<td>NGS</td>
<td>Next-generation sequencing</td>
</tr>
<tr>
<td>NIC</td>
<td>National Identification Cards</td>
</tr>
<tr>
<td>P</td>
<td>Probability</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>PD</td>
<td>Power of Discrimination</td>
</tr>
<tr>
<td>PE</td>
<td>Power of Exclusion</td>
</tr>
<tr>
<td>PIC</td>
<td>Polymorphism Information Content</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------------------------------</td>
</tr>
<tr>
<td>RLFPP</td>
<td>Restriction Fragment Length Polymorphism</td>
</tr>
<tr>
<td>RMP</td>
<td>Royal Malaysia Police</td>
</tr>
<tr>
<td>rpm</td>
<td>Revolution per minute</td>
</tr>
<tr>
<td>SNP</td>
<td>Single Nucleotide Polymorphisms</td>
</tr>
<tr>
<td>STR</td>
<td>Short Tandem Repeat</td>
</tr>
<tr>
<td>UK</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>US</td>
<td>United States</td>
</tr>
<tr>
<td>UTM</td>
<td>Universiti Teknologi Malaysia</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>VNTR</td>
<td>Variable Number of Tandem Repeats</td>
</tr>
</tbody>
</table>
**LIST OF SYMBOLS**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>µL</td>
<td>Microliter</td>
</tr>
<tr>
<td>mL</td>
<td>Mililiter</td>
</tr>
<tr>
<td>mm</td>
<td>Milimeter</td>
</tr>
<tr>
<td>°C</td>
<td>Degree Celcius</td>
</tr>
<tr>
<td>cm</td>
<td>Centimeter</td>
</tr>
<tr>
<td>kV</td>
<td>Kilovolt</td>
</tr>
</tbody>
</table>
## LIST OF APPENDICES

<table>
<thead>
<tr>
<th>APPENDIX</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendix 1</td>
<td>DNA Profile Collection Consent Form</td>
<td>126</td>
</tr>
<tr>
<td>Appendix 2</td>
<td>DNA Profile from DNA Databank Division (D13)</td>
<td>128</td>
</tr>
<tr>
<td>Appendix 3</td>
<td>Globalfiler™ Express PCR Amplification Kit loci and alleles included in the allelic ladder</td>
<td>129</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

1.1 Research Background

Deoxyribonucleic acids (DNA) are polypeptides that carry the hereditary materials in all organisms (Snustad and Micheal, 2012). It has been indicated that the diploid nuclear DNA is packaged in 44 autosomal and 2 sex chromosomes, with 23 chromosomes inherited from mother and father, individually (Buckleton et al., 2014). Datta and Datta (2012) indicated that this double-stranded macromolecule remains as one of the most powerful means for forensic human identification, whereby no two individuals would have the same DNA profile. In addition, DNA analysis is also useful for ascertaining biological kinship, as well as for criminalistics purposes (Goodwin, 2007). In this context, retrieving sufficient amount of DNA is a determinant factor for permitting suitable DNA analysis/profiling.

Being the examination of DNA at specific loci, DNA profiling approach is currently focusing on Short Tandem Repeats (STRs) that are recommended for human identification following vigorous validation studies (Romeika and Yan, 2013). This is due to the fact that they are polymorphic or hypervariable, exhibiting considerable variations among the different individuals (Romeika and Yan, 2013), resulting in unique DNA profiles. The technique used in DNA profiling is known as Multiplex Polymerase Chain Reaction (MPCR), in which DNA templates are amplified to make millions of copies of DNA at specific loci of human genome (Budowle et al., 1991).

While analysis on 16 loci (amelogenin and 15 others autosomal STRs loci) for DNA profiling had been popularly utilized in the past, a set of highly polymorphic 24 STRs loci for human individual identification has been recently suggested for forensic practical caseworks (Gettings et al., 2015). This is due to its
enhanced discrimination power following the use of 24 STRs loci (Flores et al., 2014; Martin et al., 2014). The commercial six-dye GlobalFiler™ Express PCR Amplification kit (21 autosomal STRs and 3 gender determination loci) has been successfully developed by Thermo Fisher Scientific (Wang et al., 2015). The autosomal STR loci are D3S1358, vWA, D16S539, CSF1PO, TPOX, D8S1179, D21S11, D18S51, D2S441, D19S433, TH01, FGA, D22S1045, D5S818, D13S317, D7S820, SE33, D10S1248, D1S1656, D12S391, D2S1338. On the other hand, the three gender determination loci are amelogenin, Y Indel and DYS391.

Taking into account that in DNA profiling only selected loci are analysed, the identification results obtained must be supported by robust statistical evidence. In this context, review of literature reveals variations in allele frequencies and forensic statistical parameters (e.g. power of discrimination and power of exclusion) among the different populations (Zhang et al., 2015; Ossowski et al., 2017; Adnan et al., 2018). Therefore, to substantiate the use of these 21 (excluding amelogenin, Y Indel and DYS391) loci for human identification across the varying populations, solid empirical studies those that fulfil the Daubert’s Standard for admissibility of forensic evidence must be undertaken. While various population genetics studies on the 21 STRs loci (excluding amelogenin, Y Indel and DYS391) have been reported (Park et al., 2016; Lili et al., 2017; Fujii et al., 2015), similar study focusing on the diverse Malaysian populations remains unreported, so far.

Being a multi-racial country, the population of Malaysia (32.4 million until mid-2018) is made up of Malays (69.1%), Chinese (23.0%), Indians (6.9%) and other minority ethnics (1%) (Department of Statistic Malaysia, 2018). Omar Khan (2009) briefly described the Malays, Chinese and Indians in Peninsular Malaysia as detailed below. The Malays comprise mainly Minangkabau, Bugis, Javanese as well as other groups from Indonesia, and they speak Malays and/or Indonesian languages. As for the Malaysian Chinese, they are originally from the southern part of China (mostly the provinces of Fujian and Guangdong) that arrived in Peninsular Malaysia during the 15th and mid-20th centuries. On the other hand, the Malaysian Indians are originally from the southern part of India (Madras) and majority of them are Tamils; various other Indian groups like Malayis, Telegu, Punjabis and Gujaratis are also
present (Sagoo, 2006). Therefore, it can be construed that the Malaysian population, even though smaller in size in comparison to many other populations, is diverse, necessitating the urgency of undertaking population genetic studies pertaining to the use of suitable DNA markers for human identification within the population.

1.2 Problem Statement

For the last few decades, 15 STRs loci (excluding amelogenin) were utilized for human identification in forensic cases in Malaysia with the combined probabilities of identity for the 15 STRs loci (excluding amelogenin) being approximately $2.6 \times 10^{-17}$, $7.0 \times 10^{-16}$ and $3.6 \times 10^{-17}$ for the Malays, Chinese and Indians, respectively (Seah et al., 2003). Because the use of 21 STRs loci (amelogenin, Y Indel and DYS391 were excluded) has been gaining popularity throughout the world, attributable to better discrimination ability, the DNA Databank Division (D13) of the Royal Malaysia Police (RMP) has adopted such approach for forensic intelligence since 2016 (Senior Assistant Commissioner of Police Dato’ Hussein Omar Khan, personal communication). The fact that previous population genetic studies revealed that the forensic statistical parameters varied among the different populations (Zhang et al., 2016; Alsafiah et al., 2017; Adnan., 2018) and since such data for the diverse Malaysian populations remain unreported, the robustness and evidential values of DNA evidence following the use of 21 STRs loci maybe in question in the court of law. Therefore, this present research that analysed 570 DNA profiles (190 each for Malaysian Malays, Chinese and Indians) maintained by the DNA Databank Division (D13), RMP for providing statistically evidence to support the use of the 21 STRs loci (excluding amelogenin, Y Indel and DYS391) in Malaysian perspective, appears forensically imperative. Figure 1.1 represents the conceptual framework of this present research.
Statistical analysis of population using autosomal STR loci

15 STRs Loci
(AmpF/STR® Identifiler® Direct PCR Amplification kit)

21 STRs Loci
(GlobalFiler™ Express PCR Amplification kit)

Other populations e.g.
(Zhang et al., 2016;
Alsafiah et al., 2017;
Wu et al., 2017;
Adnan et al., 2018)

Allele frequencies and established forensic statistical parameters

Empirical evidence for court presentation

Admissibility of DNA evidence in the court of law

Current study

Figure 1.1 The conceptual framework of this present research
1.3 Objectives and Hypothesis

By analysing the 21 autosomal STRs loci, the objectives of this present research were:

(a) To analyse the DNA profiles of the Malaysian Malays, Chinese and Indians in order to distinguish the allele frequencies and genotype variations.

(b) To evaluate the forensic statistical parameters (Heterozygosity, Power of Discrimination, Power of Exclusion, Matching Probability and Polymorphism Information Content) for supporting its forensic application.

It was hypothesized that the forensic statistical parameters following the use of 21 autosomal STRs loci among the Malaysian populations would be comparable with that of other populations throughout the world, indicating its robustness for forensic human identification.

1.4 Scope of Research

Following the use of GlobalFiler™ Express PCR Amplification kits, the complete DNA profiles from buccal swabs of convicted Malaysian criminal (n = 570; 190 each for Malaysian Malays, Chinese and Indians) (2016 – 2017) analysed by the author and other DNA analysts, deposited in the DNA database of the DNA Databank Division (D13) of RMP, were utilized in present research. The statistical parameters analysed included the allele frequencies, exact test of Hardy-Weinberg equilibrium including Bonferroni correction, observed and expected heterozygosity, match probability, power of discrimination, power of exclusion, matching probability, polymorphism information content and neighbor-joining tree. The analyses were carried out using Genalex v6.5, Powerstats v1.2, Arlequin v3.5, Phylip v3.69 package and TreeView32 software.
REFERENCES


119


