

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 its 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. Locus SE33 merupakan locus 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 locus 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  $1.3713 \times 10^{-25}$ ,  $2.8822 \times 10^{-25}$ ,  $7.5668 \times 10^{-26}$  dan  $1.0385 \times 10^{-26}$ . 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

	<b>TITLE</b>	<b>PAGE</b>
	<b>DECLARATION</b>	<b>ii</b>
	<b>DEDICATION</b>	<b>iii</b>
	<b>ACKNOWLEDGEMENT</b>	<b>iv</b>
	<b>ABSTRACT</b>	<b>v</b>
	<b>ABSTRAK</b>	<b>vi</b>
	<b>TABLE OF CONTENTS</b>	<b>vii</b>
	<b>LIST OF TABLES</b>	<b>xi</b>
	<b>LIST OF FIGURES</b>	<b>xiv</b>
	<b>LIST OF ABBREVIATIONS</b>	<b>xv</b>
	<b>LIST OF SYMBOLS</b>	<b>xvii</b>
	<b>LIST OF APPENDICES</b>	<b>xviii</b>
<b>CHAPTER 1</b>	<b>INTRODUCTION</b>	<b>1</b>
1.1	Research Background	1
1.2	Problem Statement	3
1.3	Objectives and Hypothesis	5
1.4	Scope of Research	5
1.5	Significance of Study	6
<b>CHAPTER 2</b>	<b>LITERATURE REVIEW</b>	<b>7</b>
2.1	Deoxyribonucleic Acid (DNA)	7
2.2	DNA in Forensic Science	8
2.3	Molecular Markers for Human Identification	9
2.3.1	Restriction Fragment Length Polymorphisms (RFLP)	9
2.3.2	Short Tandem Repeat (STR)	10
2.3.3	Single Nucleotide Polymorphism (SNP)	12
2.4	The Evolution of STRs Multiplexes	12

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
<b>CHAPTER 3</b>	<b>RESEARCH METHODOLOGY</b>	<b>23</b>
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 Heterozygosity (Ho), Expected Heterozygosity (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

<b>CHAPTER 4</b>	<b>RESULTS AND DISCUSSIONS</b>	<b>35</b>
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
<b>CHAPTER 5</b>	<b>CONCLUSION AND RECOMMENDATIONS</b>	<b>115</b>
5.1	Conclusion	115



5.2	Recommendations and Suggested Future Studies	116
-----	--	-----

<b>REFERENCES</b>		<b>118</b>
-------------------	--	------------

## LIST OF TABLES

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

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.	88
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.	89
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.	91
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.	92
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.	93
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.	94
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.	95
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.	96
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.	97
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.	98
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.	99
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.	101
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.	101

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.	102
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.	103
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.	105
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.	105
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.	109

## LIST OF FIGURES

<b>FIGURE NO.</b>	<b>TITLE</b>	<b>PAGE</b>
Figure 1.1	The conceptual framework of this present research	4
Figure 3.1	Overall flow chart of this research	27
Figure 4.1	Representative electropherogram for the DNA profile of an individual analysed in this present research.	81
Figure 4.2	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).	114

## LIST OF ABBREVIATIONS

ABI	-	Applied Biosystems
B.C	-	Before century
BDDFM	-	'Bahagian Databank DNA Forensik Malaysia'
bp	-	Base pairs
CMP	-	Combined Match Probability
CPD	-	Combined Power of Discrimination
CPE	-	Combined Power of Exclusion
DNA	-	Deoxyribonucleic Acid
DNA BASES		
A	-	Adenine
T	-	Thymine
C	-	Cytosine
G	-	Guanine
FDDM	-	Forensic DNA Databank Malaysia
FSS	-	Forensic Science Service
He	-	Expected Heterozygosity
Ho	-	Observed Heterozygosity
HWE	-	Hardy-Weinberg Equilibrium
i.e.	-	In Essence
MP	-	Match Probability
MPCR	-	Multiplex Polymerase Chain Reaction
MSDS	-	Material Safety Data Sheet
mtDNA	-	Mitochondrial DNA
NGS	-	Next-generation sequencing
NIC	-	National Identification Cards
P	-	Probability
PCR	-	Polymerase Chain Reaction
PD	-	Power of Discrimination
PE	-	Power of Exclusion
PIC	-	Polymorphism Information Content

RLFP	-	Restriction Fragment Length Polymorphism
RMP	-	Royal Malaysia Police
rpm	-	Revolution per minute
SNP	-	Single Nucleotide Polymorphisms
STR	-	Short Tandem Repeat
UK	-	United Kingdom
US	-	United States
UTM	-	Universiti Teknologi Malaysia
UV	-	Ultraviolet
VNTR	-	Variable Number of Tandem Repeats

## LIST OF SYMBOLS

$\mu\text{L}$	-	Microliter
mL	-	Mililiter
mm	-	Milimeter
$^{\circ}\text{C}$	-	Degree Celcius
cm	-	Centimeter
kV	-	Kilovolt



## LIST OF APPENDICES

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

# 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 focussing 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 15<sup>th</sup> and mid-20<sup>th</sup> 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.

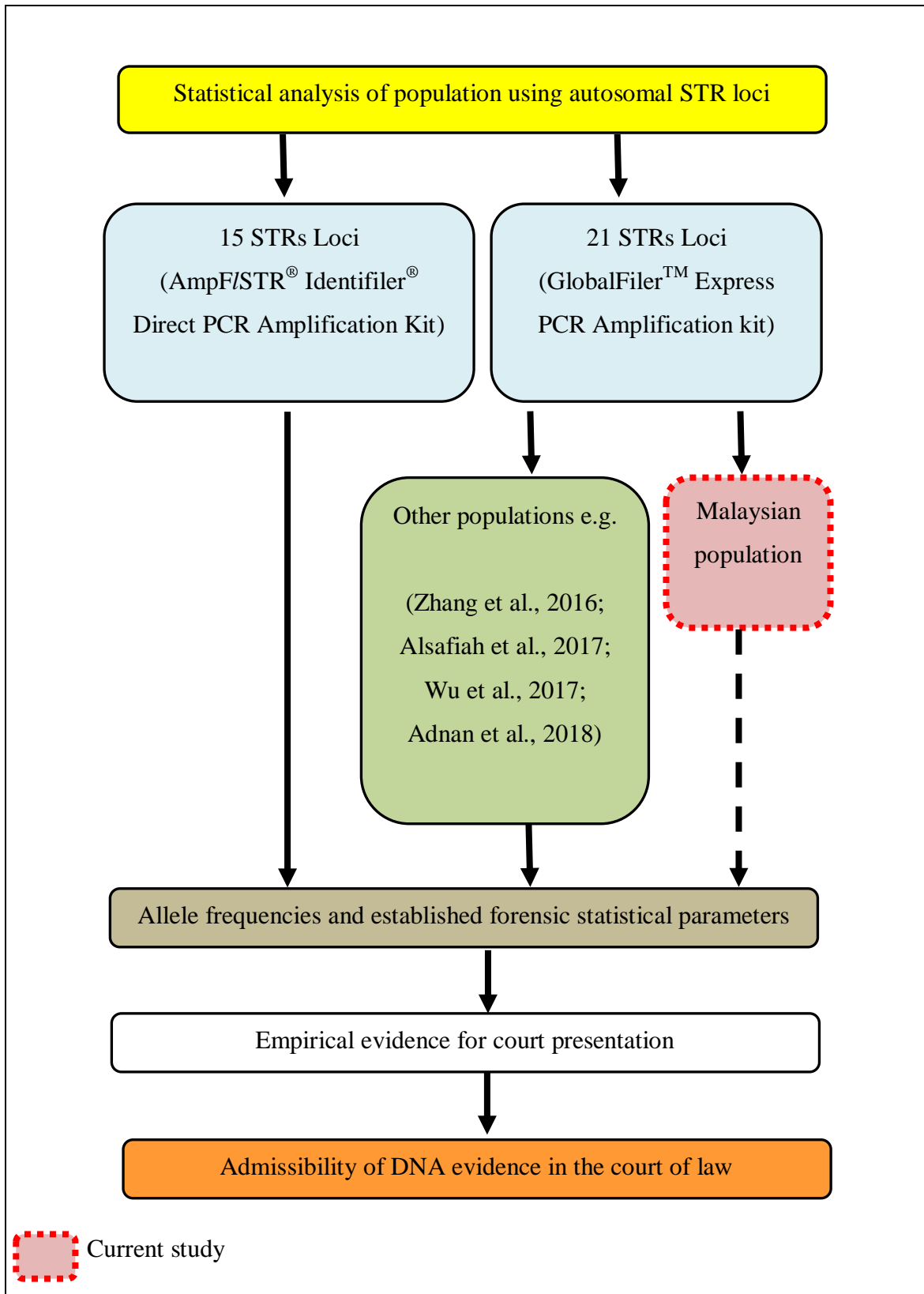


Figure 1.1 The conceptual framework of this present research

### **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

- Adnan, A., Zhan, X., Kasim, K., Rakha, A. and Xin, X.J. (2018) Population data and phylogenetic structure of Han population from Jiangsu province of China on GlobalFiler STR loci. *Forensic Science International: Genetics*, 132(5), 1301-1304. doi: 10.1007/s00414-018-1815-7.
- Alsafiah, H.M., Goodwin, W.H., Hadi, S., Alshaikhi, M.A. and Wepeba, P.cP. (2017) Population genetics data for 21 autosomal STR loci for the Saudi Arabian population using the GlobalFiler® PCR amplification kit, *Forensic Science International: Genetics*, 31, 59 – 61. doi: 10.1016/j.fsigen.2017.09.014.
- Alwi, (2016) Genetic Variations of Nine Malay Sub-Ethnic Groups in Peninsular Malaysia Using Autosomal And Y-Chromosome Short Tandem Repeats (STRs) Analysis. Master Thesis, Universiti Sains Malaysia, Kubang Kerian.
- Almeida, C., Ribeiro, T., Oliveira, A. R., Porto, M. J., Santos, J. C., Dias, D., & Dario, P. (2015). Population data of the GlobalFiler® Express loci in South Portuguese population. *Forensic Science International: Genetics*, 19, 39-41.
- Ang, H. C., Sornarajah, R., Lim, S.E.S., Syn, C.K.C., Tan-Siew, W. F., Chow S. T. and Bruce Budowle (2005) STR data for the 13 CODIS loci in Singapore Malays. *Forensic Science International: Genetics*, 148(2-3), 243-5. doi: 10.1016/j.forsciint.2004.05.009.
- Andaya, L. Y. (2001) The search for the “origins” of Melayu. *Journal of Southeast Asian Study*, 32, 1–16.
- Balamurugan, K., Kanthimathi, S., Vijaya, M., Suhasini, G., Duncan, G., Tracey, M., & Budowle, B. (2010). Genetic variation of 15 autosomal microsatellite loci in a Tamil population from Tamil Nadu, Southern India. *Legal Medicine*, 12(6), 320-323.
- Boonderm, N., Suriyanratakorn, D., Sangpueng, S., Onthong, N., Nettekul, A., W. and Waiyawuth (2017) Population Genetic Data Of 21 STR Markers in Thais of Southern Border Provinces of Thailand, *Forensic Science International: Genetics*, 523-525. doi: 10.1016/j.fsigs.2017.09.205

- Buckleton, J. S., Triggs, C.M, Walsh, S.J. (2004) *Forensic DNA Evidence Interpretation*. Boca Raton, Florida: CRC Press.
- Budowle, B., Giusti, A.M., Wayne, J.S., Baechtel, F.S., Fourney, R.M., Adams, D.E., Presley, L.A., Deadman, H.A. and Monson, K.L. (1991) Fixed bin analysis for statistical evaluation of continuous distributions of allelic data from VNTR loci for use in forensic comparisons. *American Journal of Human Genetics*, 48, 841–855.
- Budowle, B., Moretti, T. R., Keys, K. M., Koons, B. W., & Smerick, J. B. (1997). Validation studies of the CTT STR multiplex system. *Journal of Forensic Science*, 42(4), 701-707.
- Butler, J. M. (2005) *Forensic DNA typing: biology, technology, and genetics of STRs markers*. Second edn. USA.
- Butler, J. M. (2010). *Fundamentals of Forensic DNA typing*. San Diego: Academic Press.
- Butler, J.M. (2012). *Advance Topics in Forensic DNA typing: Interpretation*. San Diego: Academic Press.
- Botstein, D., White, R.L., Skolnick, M. and Davis, R.W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms, *American Journal of Human Genetics*, 32, 314–331.
- Carolyn, R., Hill David, L., Duewer, Margaret, C., Kline, Michael, D., Coble and Butler, J.M. (2012) U.S. population data for 29 autosomal STR loci, *Forensic Science International: Genetics*, 7(3), 82 – 83 doi: 10.1016/j.fsigen.2012.12.004
- Choi, E. J., Park, K. W., Lee, Y. H., Nam, Y. H., Suren, G., Ganbold, U., ... & Kim, W. (2017). Forensic and population genetic analyses of the GlobalFiler STR loci in the Mongolian population. *Genes & Genomics*, 39(4), 423-431.
- Coudray, C., Calderon, R., Guitard, E., Ambrosio, B., González-Martín, A. and Dugoujon, J. M. (2007) Allele frequencies of 15 tetrameric short tandem repeats (STRs) in Andalusians from Huelva (Spain), *Forensic Science International: Genetics*, 168(2-3), 21-24. doi: 10.1016/j.forsciint.2007.01.016.
- Datta, P., & Datta, S. (2012) Role of deoxyribonucleic acid technology in forensic dentistry. *Journal of Forensic Dental Sciences*, 4(1), 42. doi: 10.4103/0975-1475.99165



- Edinur, H. A., Zafarina, Z., Spínola, H., Nurhaslindawaty, A.R., Panneerchelvam, S. and Norazmi, M. N. (2009) HLA polymorphism in six Malay subethnic groups in Malaysia, *Human immunology*, 70(7), 518-26. doi: 10.1016/j.humimm.2009.04.003.
- Excoffier, L. and Lischer, H.E.L. (2010) Arlequin suite version 3.5: A new series of programs to perform population genetics analyses under Linux and Windows, *Molecular Ecology Resources*, 10, 564-567. doi: 10.1111/j.1755-0998.2010.02847.x
- Felsenstein, J. (2005). PHYLIP version 3.6. Software package, Department of Genome Sciences, University of Washington, Seattle, USA.
- Flores, S., Sun, J., King, J. and Budowle, B. (2014) Internal validation of the GlobalFiler™ Express PCR Amplification Kit for the direct amplification of reference DNA samples on a high-throughput automated workflow. *Forensic Science International: Genetics*, 10, 33–39. doi:10.1016/j.fsigen.2014.01.005.
- Fujii, K., Watahiki, H., Mita, Y., Iwashima, Y., Kitayama, T., Nakahara, H., Mizuno, N. and Sekiguchi, K. (2015) Allele frequencies for 21 autosomal short tandem repeat loci obtained using GlobalFiler in a sample of 1501 individuals from the Japanese population. *International Journal of Legal Medicine* (Tokyo), 17(5), 306-8. doi: 10.1016/j.legalmed.2015.08.007.
- Foreman, L. A., & Evett, I. W. (2001). *Statistical analyses to support forensic interpretation for a new 10-locus STR profiling system. International Journal of Legal Medicine*, 114, 147e155.
- Gettings, K.B., Aponte, R.A., Vallone, P.M. and Butler, J.M. (2015) STR allele sequence variation: current knowledge and future issues *Forensic Science International, Genetics*, 18, 118–130. doi:10.1016/j.fsigen. 2015.06.005.
- Gomes, V., Sánchez-Diz, P., Alves, C., Gomes, I., Amorim, A., Carracedo, A. and Gusmão, L. (2009) Population data defined by 15 autosomal STR loci in Karamoja population (Uganda) using AmpF/STR Identifiler kit. *Forensic Science International: Genetics*, 3(2), 55-58. doi: 10.1016/j.fsigen.2008.06.005.
- González, B., Velázquez, J. A., Briones, L., Chavarría, J. R., Lopez, E., Villalobos, H. (2015) Population data of 24 STRs in Mexican-Mestizo population from Monterrey, Nuevo Leon (Northeast, Mexico) based on Powerplex® Fusion

- and GlobalFiler® kits. *Forensic Science International: Genetics*, 21, 15-17. doi: 10.1016/j.fsigen.2015.12.004
- Goodwin, W., Linacre, A. and Hadi, S. (2007) *An Introduction to Forensic Genetics*. First edn. England: Wiley.
- Government of Malaysia. (2018) *Current Population Estimates, Malaysia, 2017-2018*: Department of Statistics Malaysia. [WWW Document] (31/07/2018). Available online at <https://www.dosm.gov.my/v1/index.php?r=column/pdfPrev&id=c1pqTnFjb29HSnNYNUpiTmNWZHArz09>.
- GlobalFiler® kits. *Forensic Science International: Genetics*, 21, 15-17. doi: 10.1016/j.fsigen.2015.12.004.
- Grossblatt, N. (1997). *The Evaluation of Forensic DNA Evidence*. 2nd edn. Washington: National Academy Press.
- Grossblatt, N. (1997). *DNA Technology in Forensic Science*. 3rd edn. Washington: National Academy Press.
- Gunn, A. (2009). *Essential forensic biology*. 2nd edn. UK: Wiley.
- Hamid, I. (1991) *Masyarakat dan budaya Melayu*, 2nd ed. Kuala Lumpur, Dewan Bahasa dan Pustaka, Kementerian pendidikan Malaysia.
- Hergoualc'h, M. J. (2002) *The Malay Peninsula: Crossroads of the Maritime Silk Road*. Danvers: Brill.
- Hussein, T., Ibrahim, S., Alias, R. and Jibril, J. (2007) *Malaysia negara kita (Our country Malaysia)* 2nd edn. Kuala Lumpur: MDC Sdn Bhd.
- Hussein bin Omar Khan (2009) *Inferring Ethnic Origin from STR DNA Profiles in the Malaysian Population*. Master Thesis, University of Teesside, Middlesbrough.
- Janis, E. W., David, J.C. & Goncalo, R.A. (2005) A note on exact test of Hardy-Weinberg Equilibrium. *The American Journal of Human Genetics*, 76(5),887-893.
- Jeffreys, A.J. (2005) Genetic fingerprinting, *Nature Medicine*, 11(10), 1035-9. doi: 10.1038/nm1005-1035.
- Jobling, M, Gill, P. (2004) Encoded evidence: DNA in forensic analysis, *Nature Reviews Genetic*. 6(3), 739–751.
- Jones, D. A. (1972). Blood samples: probability of discrimination. *Journal of Forensic Science Society*, 12, 355e359.

- Kashyap, V. K.1., Sarkar, N. and Trivedi R. (2002) Allele frequencies for STR loci of the Powerplex 16 multiplex system in five endogamous populations of India. *Forensic Science International: Genetics*, 126(2), 178-86.
- Kobilinsky, L., Liotti, T.F. & Sweat, J.O. (2005). *DNA: Forensic and legal applications*. New Jersey: John Wiley & Sons.
- Lancia, M., Coletti, A., G.Margiotta, Lottanti, L., Carnevali, E. and Bacci, M. (2006) Allele frequencies of 15 STR loci in an Italian population, International Congress Series, 1288, 340-342. doi:10.1016/j.ics.2005.08.019.
- Li, X., Li, L., Wang, Q., Zhang , J., Ge, W., Bai, R., Yu, X. and Shi, M. (2017) Population genetic analysis of the Globalfiler STR loci in 3032 individuals from the Altay Han population of Xinjiang in northwest China, *International Journal of Legal Medicine*, 132(1), 141 – 143. doi: 10.1007/s00414-017-1641-3.
- Martin, P., Simon, L.F., Luque, G., Farfan, M.J. and Alonso, A. (2014) Improving DNA data exchange: validation studies on a single 6dye STR kit with 24 loci. *Forensic Science International: Genetics*, 13, 68–78. doi:10.1016/j.fsigen.2014.07.002
- Maruyama, S., Minaguchi, K., Takezaki, N. and Nambiar, P. (2008) Population data on 15 STR loci using AmpF/STR Identifiler kit in a Malay population living in and around Kuala Lumpur, Malaysia, *International Journal of Legal Medicine*, 10(3), 160 – 162. doi: 10.1016/j.legalmed.2007.11.002.
- Moysés, C. B., Tsutsumida, W. M., Raimann, P. E., Motta, C. H, Nogueira, T.L., Dos Santos, O. C, Figueiredo, B. B., Mishima, T. F, Cândido, I. M., Oliveira Godinho, N. M., Beltrami, L.S., Lopes, R. K., Guidolin, A. F., Mantovani, A., Dos Santos, S. M., de Souza, C. A. and Gusmão, L. (2016) Population data of the 21 autosomal STRs included in the GlobalFiler® kits in population samples from five Brazilian regions, *Forensic Science International: Genetics*, 26, 28 – 30. doi: 10.1016/j.fsigen.2016.10.017.
- Motro, U., Adelman, R., Davidson, A., Gast, A., Hermon, D., Shpitzen, M., Zamir, A. and Freund, M. (2002) Allele frequencies of nine STR loci of Jewish and Arab populations in Israel, *International Journal of Legal Medicine*, 116(3), 184 – 186 doi: 10.1007/s004140100251.
- Ossowski, A., Diepenbroek, M., Szargut, M., Zielińska, G., Jędrzejczyk, M., Berenti, J. and Jacewicz, R. (2017) Population analysis and forensic evaluation of 21

- autosomal loci included in GlobalFiler™ PCR Kit in Poland. *Forensic Science International: Genetics*, 38 – 39. doi: 10.1016/j.fsigen.2017.05.003.
- Panneerchelvam, S. and Norazmi, M. N. (2003) Forensic DNA profiling and Database. *Malaysian Journal of Medical Sciences*, 10(2), 20 – 26.
- Park, H.C., Kim, K., Nam, Y., Park, J., Lee, J., Lee, H., Kwon, H., Jin, H., Kim, W., Kim, W. and Lim, S. (2016) Population genetic study for 24 STR loci and Y indel (GlobalFiler™ PCR Amplification kit and PowerPlex® Fusion system) in 1000 Korean individuals. *International Journal of Legal Medicine* (Tokyo), 53-7. doi: 10.1016/j.legalmed.2016.06.003.
- Peakall, R. and Smouse, P.E. (2012) GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics*, 28(19), 2537–2539. doi: 10.1093/bioinformatics/bts460.
- Rakha, A., Bing, Y., Hadi, L. and Sheng-bin. (2009) Population genetic data on 15 autosomal STRs in a Pakistani population sample. *International Journal of Legal Medicine*, 11(6), 305-307. doi:10.1016/j.legalmed.2009.08.001.
- Ristow, P. G., Cloete, K. W., D'Amato, M. E. (2016) GlobalFiler® Express DNA amplification kit in South Africa: Extracting the past from the present, *Forensic Science International: Genetics*, 24, 194-201. doi: 10.1016/j.fsigen.2016.07.007.
- Romeika, J. M., Yan, F. (2013) Recent Advances in Forensic DNA Analysis. *Journal Forensic Research*, S12: 001. doi: 10.4172/2157-7145.S12-001.
- Sagoo, K. (2006) Colonial Construction of Malayness. The Influence of Population Size and Composition. 5th East-West Center International Graduate Student Conference. 16 – 18 February 2006. Honolulu, Hawaii USA: 1 – 13.
- Sainuddin, S. (2003) Titas tamadun Melayu (Malay civilization Titas). Perak: Quantum Books.
- Seah, L.H., Jeevan, N.H., Othman, M.I., Jaya, P., Ooi, Y.S., Wong, P.C. and Kee, S.S. (2003) STR Data for the AmpFISTR Identifiler loci in three ethnic groups (Malay, Chinese, Indian) of the Malaysian population. *Forensic Science International: Genetics*, 138(1-3), 134-7. doi: 10.1016/j.forsciint.2003.09.005.
- Sensabaugh, G. (1982). Biochemical markers of individuality. In R. Saferstein (Ed.), *Forensic Science Handbook* (pp. 338e415). New York: Prentice-Hall.

- Stephenson, F. H. (2016) *Calculations for Molecular Biology and Biotechnology*. 2nd edn. USA: Academic Press.
- Snustad, D. P. and Simmons, M. J. (2011). *Principle of Genetics*. 6th ed: Wiley.
- Suhr, D (2005) Principal component analysis vs. exploratory factor analysis. SAS Conference Proceedings of the SAS: Users Group International 30. April 10–13. Philadelphia, PA. Cary, NC: SAS Institute Inc, 203–230.
- Thermo Fisher Scientific. (2016) *Globalfiler™ Express PCR Amplification Kit User Guide*. Applied Biosystems.
- Untoro E., Atmadja, D. S., Pu, C. E. and Wu, F. C. (2009) Allele frequency of CODIS 13 in Indonesian population, *International Journal of Legal Medicine*, 11(1), S203 – 205. doi: 10.1016/j.legalmed.2009.01.007.
- Wang, D.Y., Gopinath, S., Lagacé, R.E., Norona, W., Hennessy, L.K., Short, M.L. and Mulero, J.J. (2015) Developmental validation of the GlobalFiler® Express PCR Amplification Kit: a 6-dye multiplex assay for the direct amplification of reference samples. *Forensic Science International: Genetics*, 19, 148–155. doi:10.1016/j.fsigen.2015.07.013.
- White, P.C. (2005) *Crime Scene to Court: The Essentials of Forensic Science*. Second Edition. UK: Royal Society of Chemistry.
- William, C. H. (2005) *Reproductive Biology and Phylogeny of Chondrichthyes Sharks, Batoids and Chimaeras*. 3rd edn. USA: Science Publishers, Inc.
- Wu, L., Pei, B., Ran, P. and Song, X. (2017). Population genetic analysis of Xiamen Han population on 21 short tandem repeat loci. *International Journal of Legal Medicine (Tokyo)*, 41-44. doi: 10.1016/j.legalmed.2017.03.002.
- Zhang, H., Shuping, Y., Wei, G., Ren, B., Liwen, P., Teng, M., Mingying, X., Jin, L., Liming, L. and Shilin, L. (2016) Population genetic analysis of the GlobalFiler STR loci in 748 individuals from the Kazakh population of Xinjiang in northwest China. *International Journal of Legal Medicine*, 130(5), 1187-9. doi: 10.1007/s00414-016-1319-2.
- Zhang, X., Du, L., Huang, 3., Gu, T., Hu, L. and Nie, S. (2017) Genetic variation of 20 autosomal STR loci in three ethnic groups (Zhuang, Dai and Hani) in the Yunnan province of southwestern China, *Forensic Science International: Genetics*, 31, 41-42. doi: 10.1016/j.fsigen.2017.06.005.
- Zhivotovsky, L. A., Malyarchuk, B. A., Derenko, M. V., Wozniak, M. and Grzybowski, T. (2009) Developing STR databases on structured populations: