ALLEL FREQUENCY DISTRIBUTION OF HYPERVARIABLE VNTR D17S5 LOCUS IN CHINESE COMMUNITY IN UNIMAS

Jessie Ann Sim Chiew Ling

Bachelor of Science with Honours (Resource Biotechnology) 2006
ALLEL FREQUENCY DISTRIBUTION OF HYPERVARIABLE VNTR D17S5 LOCUS IN CHINESE COMMUNITY IN UNIMAS

JESSIE ANN SIM CHIEW LING

This project is submitted in partial fulfillment of the requirements for the degree of Bachelor of Science with Honours (Resource Biotechnology)

Faculty of Resource Science and Technology
UNIVERSITI MALAYSIA SARAWAK

2006
ACKNOWLEDGEMENTS

This project was performed in partial fulfillment for the Degree of Bachelor of Science with Honors. First and foremost, I would like to convey my deepest gratitude and appreciation to my respected supervisor, Dr. Hairul Azman Roslan for his continuous guidance, encouragement and patience throughout the whole duration of conducting this study. I would also like to thank all the lab assistants especially Mr. Azis Ajim and all the postgraduate students namely Mr. Tan Sia Hong, Ms. Patricia Chong, Ms. Yasotha and Mr. Ang Chung Huap for all their relentless guidance, help and support while conducting this study.

Besides that, I would also like to express my appreciation to my fellow coursemates for their suggestions and help that contributed to my completing this project. Last but not least, to my family who has been providing me with all their moral support all along. I am highly grateful for their love and encouragement.
ALLELE FREQUENCY DISTRIBUTION OF HYPERVARIABLE VNTR D17S5 LOCUS
OF CHINESE COMMUNITY IN UNIMAS

Jessie Ann Sim Chiew Ling
Resource Biotechnology Program
Faculty of Resource Science and Technology
University of Malaysia Sarawak

ABSTRACT

The high polymorphism characteristic of variable number of tandem repeats (VNTRs) have spurred studies based on it to be carried out and its applications covers a wide range which varies from genetic identification to population genetics and forensic medicine. For this study, allele frequencies at the hypervariable VNTR D17S5 locus were determined from the Chinese population sample in UNIMAS via the polymerase chain reaction (PCR) method. 50 samples were analyzed out of the 75 samples collected and this yielded sufficient results for statistical evaluations. The distributions of the D17S5 alleles for the Chinese population in UNIMAS were the highest in allele number 2 and lowest in allele number 5.

Keywords: VNTR, D17S5, polymerase chain reaction (PCR), allele frequency.

ABSTRAK

Ciri unik yang diperolehi oleh "variable number of tandem repeats" (VNTRs) iaitu polimorfisma yang tinggi telah menggalakkan kajian dilakukan berasaskaninya dan kegunaannya merangkumi skop yang luas umpsamanya identifikasi genetik, genetik populasi dan bidang forensik perubatan. Dalam kajian ini, frekuensi alel pada lokus VNTR D17S5 telah dikenalpasti dari sampel populasi Cina di UNIMAS menggunakan kaedah "polymerase chain reaction" (PCR). 50 sampel telah dianalisa dari jumlah 75 sampel yang dikumpul dan ini telah menghasilkan keputusan yang memadai untuk analisa statistik. Penumpuan alel D17S5 adalah paling tinggi di alel nombor 2 dan paling rendah di alel nombor 5 untuk populasi Cina di UNIMAS.

Kata kunci: VNTR, D17S5, polymerase chain reaction (PCR), frekuensi alel.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CONTENTS</th>
<th>PAGE NUMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>i</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>ABSTRAK</td>
<td>ii</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>iii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>ix</td>
</tr>
<tr>
<td>CHAPTER 1 INTRODUCTION</td>
<td></td>
</tr>
<tr>
<td>1.1 Variable Number of Tandem Repeats (VNTRs)</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Characteristics of VNTR D17S5</td>
<td>2</td>
</tr>
<tr>
<td>1.3 Chinese Community in UNIMAS</td>
<td>2</td>
</tr>
<tr>
<td>1.4 Polymerase Chain Reaction (PCR)</td>
<td>3</td>
</tr>
<tr>
<td>1.5 Agarose Gel Electrophoresis (AGE)</td>
<td>5</td>
</tr>
<tr>
<td>1.6 Statistical Analyses</td>
<td>7</td>
</tr>
<tr>
<td>1.7 Objectives</td>
<td>8</td>
</tr>
<tr>
<td>1.8 Literature Reviews</td>
<td>8</td>
</tr>
<tr>
<td>CHAPTER 2 MATERIALS AND METHODOLOGIES</td>
<td>11</td>
</tr>
<tr>
<td>2.1 Sample Collection</td>
<td>11</td>
</tr>
<tr>
<td>2.2 DNA Extraction and Quantification</td>
<td>12</td>
</tr>
<tr>
<td>(Richards et al., 1993)</td>
<td></td>
</tr>
<tr>
<td>2.3 Polymerase Chain Reaction (Faykina et al., 1996)</td>
<td>13</td>
</tr>
<tr>
<td>2.4 Agarose Gel Electrophoresis (Faykina et al., 1996)</td>
<td>14</td>
</tr>
<tr>
<td>2.5 Statistical Analyses</td>
<td>15</td>
</tr>
<tr>
<td>CHAPTER 3 RESULTS AND DISCUSSION</td>
<td>16</td>
</tr>
<tr>
<td>3.1 Spectrophotometer Readings of Samples</td>
<td>16</td>
</tr>
<tr>
<td>3.2 AGE Visualization of Amplified D17S5 Alleles</td>
<td>18</td>
</tr>
<tr>
<td>3.3 Statistical Analyses</td>
<td>22</td>
</tr>
</tbody>
</table>
CHAPTER 4 CONCLUSIONS
CHAPTER 5 REFERENCES
CHAPTER 6 APPENDICES

6.1 Consent Form
6.2 Spectrophotometer Readings of Samples
6.3 Linear Regression Equations for the Molecular Weight Standard of Each Gel
  6.3.1 Gel 1 (Figure 3)
  6.3.2 Gel 2 (Figure 4)
  6.3.3 Gel 3 (Figure 5)
  6.3.4 Gel 4 (Figure 6)
  6.3.5 Gel 5 (Figure 7)
  6.3.6 Gel 6 (Figure 8)
6.4 Raw Datasets
  6.4.1 CONVERT131 Input Data
  6.4.2 Popgene Input Data
6.5 D17S5 Allele Frequencies Computed via Popgene
LIST OF TABLES

Table 1. Summary of tests conducted using the POPGENE software package for allele frequency of the VNTR D17S5 locus and their functions (Yeh et al., 1997) 7

Table 2. VNTR PCR primers (Horn et al., 1989) 13

Table 3. PCR conditions (modified from Faykina et al., 1996) 13

Table 4. Spectrophotometer readings of DNA concentration and purity in ascending order of DNA concentration 39

Table 5. Molecular weight standards for gel 1 according to the left ladder in figure 3 41

Table 6. Molecular weight standards for gel 1 according to the right ladder in figure 3 42

Table 7. Molecular weight standards for gel 2 according to the left ladder in figure 4 43

Table 8. Molecular weight standards for gel 2 according to the right ladder in figure 4 44

Table 9. Molecular weight standards for gel 3 according to the ladders in figure 5 45

Table 10. Molecular weight standards for gel 4 according to the left ladder in figure 6 46

Table 11. Molecular weight standards for gel 4 according to the right ladder in figure 6 47

Table 12. Molecular weight standards for gel 5 according to the left ladder in figure 7 48

Table 13. Molecular weight standards for gel 5 according to the right ladder in figure 7 49

Table 14. Molecular weight standards for gel 6 according to the ladder in figure 8 50
| Table 15. | Samples’ migration distance on AGE and linear regression values generated using Microsoft Excel software |
| Table 16. | Data of sample size(s) for input into Convert131 |
| Table 17. | Samples’ data converted into Popgene format according to ascending order in sizes |
| Table 18. | Allele frequencies for the D17S5 locus generated using Popgene software |
| Table 19. | Summary of genic variation and heterozygosity statistics for D17S5 locus generated using Popgene |
| Table 20. | Allele frequencies calculated manually according to the equation reported by Carr (2004) |
LIST OF FIGURES

Figure 1. Schematic representation of denaturation, annealing and extension processes in PCR amplification 4
Figure 2. Schematic representation of the amplification of the desired VNTR locus and separation of the PCR products using gel electrophoresis 5
Figure 3. AGE visualization of PCR products 18
Figure 4. AGE visualization of PCR products 19
Figure 5. AGE visualization of PCR products 20
Figure 6. AGE visualization of PCR products 20
Figure 7. AGE visualization of PCR products 21
Figure 8. AGE visualization of PCR products 21
Figure 9. Linear regression for the molecular weight standards of gel 1 according to the left ladder in figure 3 41
Figure 10. Linear regression for the molecular weight standards of gel 1 according to the right ladder in figure 3 42
Figure 11. Linear regression for the molecular weight standards of gel 2 according to the left ladder in figure 4 43
Figure 12. Linear regression for the molecular weight standards of gel 2 according to the right ladder in figure 4 44
Figure 13. Linear regression for the molecular weight standards of gel 3 according to the ladders in figure 5 45
Figure 14. Linear regression for the molecular weight standards of gel 4 according to the left ladder in figure 6 46
Figure 15. Linear regression for the molecular weight standards of gel 4 according to the right ladder in figure 6 47
Figure 16. Linear regression for the molecular weight standards of gel 5 according to the left ladder in figure 7 48
Figure 17. Linear regression for the molecular weight standards of gel 5 according to the right ladder in figure 7

Figure 18. Linear regression for the molecular weight standards of gel 6 according to the ladder in figure 8
## LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE</td>
<td>Agarose Gel Electrophoresis</td>
</tr>
<tr>
<td>dATP</td>
<td>2'-deoxyadenosine-5'-triphosphate</td>
</tr>
<tr>
<td>dCTP</td>
<td>2'-deoxycytidine-5'-triphosphate</td>
</tr>
<tr>
<td>dGTP</td>
<td>2'-deoxyguanosine-5'-triphosphate</td>
</tr>
<tr>
<td>dTTP</td>
<td>2'-deoxythymidine-5'-triphosphate</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic Acid</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetra Acetic Acid</td>
</tr>
<tr>
<td>FRST</td>
<td>Faculty of Resource Science and Technology</td>
</tr>
<tr>
<td>$H_{\text{exp}}$</td>
<td>Expected Heterozygosity</td>
</tr>
<tr>
<td>$H_0$</td>
<td>Observed Heterozygosity</td>
</tr>
<tr>
<td>NaOH</td>
<td>Sodium Hydroxide</td>
</tr>
<tr>
<td>PAGE</td>
<td>Polyacrylamide Gel Electrophoresis</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>STR</td>
<td>Short Tandem Repeat</td>
</tr>
<tr>
<td>TAE</td>
<td>Tris-acetate-EDTA</td>
</tr>
<tr>
<td>Tris-HCl</td>
<td>Tris-hydrochloric-acid</td>
</tr>
<tr>
<td>VNTR</td>
<td>Variable Number of Tandem Repeat</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

1.1 Variable number of tandem repeats (VNTRs)

About 30% of the human genome is made up of human tandem repeats, excluding satellite DNA, which are classified into two groups according to the size of the repeat unit and the total length of the repeat array (Tamaki and Jeffreys, 2005). The human minisatellites or also known as variable number tandem repeat (VNTR) loci have repeat units from 6 to more than 100 base pairs long depending on the locus, with arrays usually kilobases in length (Tamaki and Jeffreys, 2005). In addition to that, VNTRs vary in different individuals (Consolidated Safety Services, n.d.) and they represent a single locus in the human genome (Gasparini et al., 1991). The genotype of VNTRs is basically co-dominant whereby it means that the products of both alleles on a locus can be detected in heterozygotes (Hartl and Clark, 1997). For this study, suites of single locus alleles were obtained from the visualization of the D17S5 locus via the gel electrophoresis methods. The results obtained from these enabled the easy interpretation of the genetics of the single locus alleles and probability calculations carried out straightforwardly (Strobeck, 1993).

The unique characteristic of high polymorphism possessed by VNTRs results in a number of different alleles which makes the possibility of two discrete samples sharing the same allele low (Faykina et al., 1996). Besides that, this exact feature of the VNTRs also makes them very suitable markers for genetic identification, as individuals will regularly show different genotypes at the locus (Gasparini et al., 1991). Other applications of VNTRs also include forensic medicine, paternity testing and population genetics (Pinheiro et al., 1995).
1.2 Characteristics of VNTR D17S5

The VNTR D17S5 marker is located on chromosome number 17 (Faykina et al., 1996) and it comprise of 19 repeat units (Kijas et al., 1994). The repeat units has tandem repeats 70 base pairs in length (Faykina et al., 1996) thus the alleles range from 170 to 1430 base pairs in length (Kijas et al., 1989).

1.3 Chinese community in UNIMAS

The Chinese community in UNIMAS comes from all over Malaysia including Sabah and Sarawak and many of these individuals have mixed parentage or ancestry. As the Chinese community actually originates from China, the Chinese community comprise of few different dialects some of which are Teochew, Hokkien, Hakka, Cantonese and many others.

To date, there are no studies done based on the allele frequency distribution of the hypervariable VNTR D17S5 locus in the Chinese community in neither Malaysia nor in UNIMAS yet. Therefore, this study aimed to determine the base pair lengths of the VNTR D17S5 alleles and from there generate the allele frequency distribution of the locus via various tests as outlined in table 1 using the Popgene software.
1.4 Polymerase Chain Reaction (PCR)

Basically, the PCR method is the in vitro enzymatic amplification of a specific DNA fragment of known sequence by consecutive rounds of DNA synthesis yielding microgram quantities from a minute amount of initial substance (Reece, 2004). The DNA target sequence information is required to create two amplimers, oligonucleotide primer sequences (often 20-30 nucleotides long) that will exclusively bind to complementary DNA immediately flanking the desired target region when added to genomic DNA (Strachan, 1992). The amplimers are designed so that they can produce new DNA strands which are complementary to the individual DNA strands of the target DNA segment and which will overlap each other in the presence of DNA polymerase and DNA precursors (Strachan, 1992). The DNA polymerase for use in this study is the Taq polymerase while the DNA precursors consist of four nucleotides (dTTP, dATP, dCTP and dGTP).

The three main steps in each PCR cycle are denaturation, annealing and extension. During denaturation (Figure 1-A), the two strands of the target DNA molecule are separated into its component strands by heating to approximately 95°C (Reece, 2004). Consequently, the two target strands are then cooled to allow the oligonucleotide primers to bind to the dissociated strands during annealing at 55°C (Figure 1-B) (Consolidated Safety Services, n.d.). One of the primers will then recognize and bind to one of the target DNA strands while the other primer recognizes and binds to the other DNA strand (Reece, 2004). The primers are designed in such a way that the 3' end of each primer faces the other one, and so DNA synthesis proceeds on both strands through the region between the two primers (Reece, 2004). Finally, during extension (Figure 1-C), the solution is reheated to about 75°C to synthesize complementary copies of each strand.
It is during this that the DNA polymerase binds to the free 3'-end of each of the bound oligonucleotides and uses dNTPs to synthesize a new DNA strand in a 5' to 3' direction (Reece, 2004). The size of the amplified locus can then be determined following PCR via gel electrophoresis (Brown, 1990). Figures 1 and 2 thus illustrate the whole process of amplification of the desired VNTR locus and separation of the PCR products via gel electrophoresis.

Figure 1. Schematic representation of denaturation, annealing and extension processes in PCR amplification. (Taken from: http://genetics.nhii.gov/Basic2.html)
Figure 2. Schematic representation of the amplification of the desired VNTR locus and separation of the PCR products using gel electrophoresis. (Taken from: http://www.essentialcellbiology.com)

1.5 Agarose Gel Electrophoresis (AGE)

The deoxyribonucleic acid (DNA) consists of long, thin and negatively charged molecules (Reece, 2004) which vary in length and conformation. Thus, in order to determine the sizes of the molecules, AGE is employed to separate them on the basis of their properties. The fundamental principal behind AGE is the application of electric current through the gel to facilitate the migration of DNA molecules from their initial position at the negative electrode (cathode) toward the positive electrode or anode (Reece, 2004). The movement of DNA through the pores of the gel matrix is like reptating or snaking (Reece, 2004). The gel matrix acts like a sieve which enables small DNA fragments to move easier compared to larger DNA fragments which may get entangled within the gel pores (Reece, 2004).
After separation, the gel containing the migrated DNA fragments are then soaked in an ethidium bromide solution to allow for visualization (Reece, 2004) via the ultraviolet transilluminator. Ethidium bromide is an intercalating dye which fluoresces under ultraviolet light that integrates into the grooves of DNA (Consolidated Safety Services, n.d.) resulting in the distortion and localized unwinding of the double-helix structure (Reece, 2004). After that, the soaked gel are illuminated with light in the ultraviolet range (260-300 nm), which produces a fluorescence of ethidium bromide, and the DNA are seen on the gel as a band of fluorescence (Reece, 2004). Then, each sample is determined to be homozygous or heterozygous for the D17S5 allele by the number of bands appearing on the agarose gel (Faykina et al., 1996). As one marker has only one copy per haploid genome thus one band indicates homozygosity whereas two bands indicate heterozygosity in the diploid human genome (Faykina et al., 1996). This method is also used to determine the base pair length of the alleles on different individuals as the D17S5 locus consists of tandem repeats of 70 base pairs in length (Faykina et al., 1996) and this can be sufficiently separated using AGE.
1.6 Statistical Analyses

To calculate the frequency of homozygous and heterozygous samples in a given population, the data obtained from the number of bands appearing on the agarose and gel were used. In this study, the statistical analyses of the allele frequency for VNTR D17S5 were calculated using Microsoft Excel, Convert131 and Popgene softwares. A number of tests that comprise of the allele number, allele frequency, effective allele number, observed homozygosity and observed heterozygosity for the VNTR D17S5 locus were calculated using Popgene (Yeh et al., 1997). The functions of each test are summarized in Table 1.

<table>
<thead>
<tr>
<th>Tests</th>
<th>Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allele frequency</td>
<td>Estimate gene frequencies at the VNTR D17S5 from raw data</td>
</tr>
<tr>
<td>Allele number</td>
<td>Counts the number of alleles with nonzero frequency</td>
</tr>
<tr>
<td>Effective allele number</td>
<td>Estimates the reciprocal of homozygosity</td>
</tr>
<tr>
<td>Observed homozygosity</td>
<td>Estimate the proportion of observed homozygotes at the VNTR D17S5 loci</td>
</tr>
<tr>
<td>Observed heterozygosity</td>
<td>Estimate the ratio of observed heterozygotes at the VNTR D17S5 locus only for co-dominant markers.</td>
</tr>
</tbody>
</table>
1.7 Objectives

The objectives of this study were to:

1) amplify the hypervariable VNTR D17S5 alleles
2) determine the base pair length of the alleles of the VNTR D17S5 locus
3) compute the standard allelic frequencies and distribution of the D17S5 alleles
4) establish a genetic database based on the allele frequency distribution of the D17S5 allele of the Chinese community in UNIMAS

1.8 Literature Reviews

A number of similar researches have been done on the VNTR D17S5 locus in various populations all over the world pertaining to the techniques used via exploiting the locus and its applications. One of the researches based on that was carried out by Horn and his fellow researchers (1989) on the amplification of the pYNZ22 locus, which in this case also refers to the VNTR D17S5. More than ten alleles have been observed by Southern blotting that contained 11 repeat units, ranging from 170 base pairs to 870 base pairs with a heterozygosity of 86%. Gecz also carried out a similar study in 1991 whereby a new PCR protocol based on lower temperature and longer extension time were designed. This procedure resulted in the discovery of alleles containing 12 and 14 repeat units and this brings the total number of D17S5 alleles to 14 and observed heterozygosity to 88%.

While in 1994, Hatzaki et al. analyzed variability at D17S5 out of five VNTR loci in the Greek population using PCR amplification. 11 alleles were observed while the alleles with the highest and lowest frequencies were alleles number 4 and 6 respectively. A heterozygote
deficiency was indicated at this locus due to the failure of smaller alleles to amplify under standard conditions when present with larger alleles in heterozygosity. However, this problem was alleviated by decreasing the template DNA concentration and increasing the extension time. Following that, allele and genotype frequencies for the YNZ22 locus were determined in a population sample of the North of Portugal using the PCR amplification and nonradioactive detection by Pinheiro et al. (1995). In this study, 12 alleles were found with the highest distribution of alleles centered also on allele number 4 while allele number 7 has the lowest allele frequency. Then, in 1996, Faykina et al. conducted an analysis of four polymorphic human genetic markers by PCR. They found 12 segregating alleles that ranged from 168-938 base pairs in length.

Subsequently, genetic relationships among the Japanese, Northern Han, Hui, Uygur, Kazakh, Greek, Saudi Arabian and Italian populations based on allelic frequencies at four VNTR and one STR locus were studied by Katsuyama et al. in the year 1998. The result for the D17S5 locus showed that 14 alleles was observed in the eight human populations studied and the distribution patterns for alleles 1 and 2 varied distinctly among the eight populations observed.

In addition to all the studies carried out based on the D17S5 locus in the 20th century, Das and his co-researchers conducted one on the genetic polymorphism at four minisatellite loci (D1S80, D17S5, D19S20, and APOB) among five Indian population groups namely Konkanastha Brahmins, Marathas, Nairs, Ezhavas, and Muslims in the year 2002. The techniques utilized were PCR amplification and polyacrylamide gel electrophoresis (PAGE). It was found that the distribution of the D17S5 alleles among 345 individuals analyzed from the five population groups varied, with 13 in Marathas, 12 in Konkanasthas, 11 in Muslims, and 10 each in Nairs and
Ezhavas and from these, eight alleles were found to be common among the five population groups (Das et al., 2002).

Another research was also conducted pertaining to the application of VNTR by Kamel et al. (2002) in utilizing VNTRs as a tool of chimerism detection in allogeneic stem cell transplantation using PCR amplification. The result of this research revealed that a total of 9 alleles were reported for the YNZ22 locus with allele number 1 having the highest distribution and allele number 9 having the lowest (Kamel et al., 2002). The observed heterozygosity index was much lower than the expected and this means a heterozygote deficiency was detected. Besides that, this study concluded that using PCR for the amplification of locus was a dependable, quick and sensitive substitute to Southern hybridization (Kamel et al., 2002).

In addition to all the other studies, Tenaglia and her fellow researchers (2004) performed a study on the allele frequencies of nine STRs (CSF1PO, TPOX, THO1, F13A01, FESFPS, vWA, D16S539, D7S820 and D13S317) and two VNTRs (D1S80 and D17S5) in the Rosario population of Santa Fe Argentine. Their study yielded 11 alleles from the D17S5 locus whereby alleles number 4 and 6 has the highest and lowest frequencies respectively. There were no differences found when the statistical data from this study was compared with those of other Argentine populations and of other population studies from Central and South America.

Also in 2004, Chang et al. reported the allele frequencies and genetic data of 16 highly polymorphic loci in Koreans. There were 27 alleles observed in the D17S5 locus ranging from 1 to 43 repeat units from a sample of 598 unrelated Koreans. Allele number 1 has the highest concentration of alleles while allele number 31 has the lowest. There were no significant differences when comparing the data of this study to earlier published data of the same loci.
CHAPTER 2
MATERIALS AND METHODOLOGIES

2.1 Sample Collection

Samples of cheek cells were collected randomly from 75 unrelated Chinese UNIMAS students using sterile toothpicks. Before that, the individuals donating their DNA samples were verified of the purity of their race based on verbal communication and then requested to sign a consent form indicating their permission to allow his or her DNA to be used exclusively by the Faculty of Resource Science and Technology (FRST) for research and study purposes. The consent form and name list of the donors can be found in the Appendix section 6.1, pages 35-38.

Sample collection was carried out by the donors carefully swabbing their inner cheek walls for a few times with the blunt end of the provided sterile toothpick. Next, that end of the particular toothpick was dipped into an eppendorf tube filled with 100 μl of sterile water. These procedures were then repeated for another 5-6 times until the water in the tube becomes cloudy. The samples were then stored at -20°C before processing.
2.2 DNA Extraction and Quantification (Richards et al., 1993)

This buccal cell processing method was based on the study by Richards et al. (1993) with some modifications. The purpose of this protocol was to extract genomic DNA from the collected cheek cells samples. Firstly, 5 µl of 1M NaOH was added to the cell suspensions. This brought the mixture concentration to 50 µM. Then, the mixture was centrifuged at 13 000 rpm for 1 minute. A hole was poked through the lid of each tube with sterile needles to prevent the tube from popping open due to pressure when incubated in the waterbath. The tube was then incubated in the waterbath for 15 minutes at a temperature of 90°C, then removed from the waterbath and allowed to cool for another 15 minutes. 5.5 µl of 1M Tris-HCl pH7.5 was subsequently added to the mixture to neutralize the samples. Subsequently, the suspension was mixed until the content is even. The solution was then centrifuged at 13 000 rpm for 10 minutes. After that, the supernatant was extracted from the solution and transferred into a new 1.5 ml eppendorf tube. DNA was quantified on an Ultrospec® 1100 pro spectrophotometer (Amersham Pharmacia Biotech). DNA purity was assessed using the A260:280 ratio and DNA concentration was measured against calibration prepared from distilled water. DNA quantification was carried out by adding 5 µl of DNA sample into a quartz cuvette tube before totaling it up with 495 µl of distilled water. The cuvette was then covered with a small piece of Parafilm and inverted several times to mix thoroughly before quantification. Finally, the samples were stored at −20°C.
2.3  Polymerase Chain Reaction (Faykina et al., 1996)

This PCR method was modified from the original protocol based on Faykina et al. (1996). The reagents to be used for a 1X PCR mix will be added in the following order:

- 7.5 µl sterile water,
- 2.5 µl 10X buffer with (NH₄)₂SO₄,
- 1 µl dNTP mix,
- 1 µl D17S5 Forward primer,
- 1 µl D17S5 Reverse primer,
- 0.5 µl Taq DNA polymerase,
- 10 µl DNA and lastly, 1 drop of mineral oil (optional).

The sequences of the D17S5 Forward and Reverse primers and the PCR conditions are shown in Tables 2 and 3, respectively.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Chromosome localization</th>
<th>Sequences</th>
</tr>
</thead>
</table>
| D17S5 | 17p13 | D17S5 Forward primer: 5'-CACAGTCCTTTATTCACCG-3'  
D17S5 Reverse primer: 5'-CGAAGAGTGAAGTGCACAGG-3' |

<table>
<thead>
<tr>
<th>VNTR Locus</th>
<th>PCR Conditions (modified from Faykina et al., 1996)</th>
</tr>
</thead>
</table>
| D17S5 | Initial Denaturation: 94°C, 3 mins  
Denaturation: 94°C, 30 secs  
Annæaling: 54°C, 1 min  
Extension: 72°C, 1 min  
Final at 72°C, 10 mins  
Cycles: 35 |