UNIVERSITI SAINS MALAYSIA

Polymorphism of STRs D7S820, D13S317, and D16S539 in Dusun population

Dissertation submitted in partial fulfillment for the

Degree of Bachelor of Science (Health) in Forensic Science

Liew Chyi

School of Health Sciences
Universiti Sains Malaysia
Health Campus
16150, Kubang Kerian, Kelantan
Malaysia

2006

CERTIFICATE

This is to certify that the dissertation entitled

"Polymorphism of STRs D7S820, D13S371, and D16S539 in Dusun population"

is the bonafide record of research work done by

Ms. Liew Chyi

during the period November to April
under my supervision.

Signature of Supervisor:

Name and address of Supervisor: Mr. S. Panneerchelvam
School of Health Sciences
Universiti Sains Malaysia

Date: 11.05.06

Signature of Co Supervisor:

Name and address of Co Supervisor: Professor Dr. Norazmi Mohd. Noor
School of Health Sciences
Universiti Sains Malaysia

Date:

ACKNOWLEDGEMENTS

Considerable help has been obtained directly or indirectly from many parties while I worked out this research paper. With this opportunity I have been greatly indebted to them for their help, advice and encouragement. I wish to extend my deep appreciation to my research supervisor, Mr. Panneerchelvam, for his guidance and support at various stages of development during this study. I am grateful for his kindness spent the valuable time reviewing and provided constructive criticism to my writing. I also wish to thank my research co supervisor, Professor Dr. Norazmi Mohd. Noor for spending his time to figure out this research.

Also, I am grateful acknowledges my indebtedness to Ms. Rosniah and Ms. Linda for the excellent technical assistant, advices and suggestions in the laboratory. I am thankful to my friends who were there offered help whenever I needed. Especially, I wish to thank to Ivy Eddie, for her gift of collected buccal swab samples from her local population.

Without the encouragement and help of them this work would not have been possible. Their cooperation and generous courtesies at all times made the completion of this study a pleasant experience. Thank you.

Liew Chyi

72748

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

ABSTRACT

DNA typing of short tandem repeat (STR) loci is routinely done for DNA analysis in forensic practices. This PCR-based technique is useful for paternity and personal identification testing. STR technology evaluates the specific loci that found in nuclear DNA which are highly polymorphic in nature. These highly polymorphic markers can be typed from DNA derived from various forensic biological samples. Once the multiple locus profile is established, it intensifies the discrimination between one DNA profile and another. As the number of analyzed loci increases, the probability of a second, unrelated person, excluding identical twins, having the same profile becomes ever unlikely. Therefore, the population data on the distribution of allele frequencies are essential for the estimation of probability of certainty. It would be of benefit for the forensic community to make data on the genetic variation of selected validated loci in as many populations as possible. This study examined the population frequency and genetic diversity of three autosomal STR loci, D7S820, D13S371, and D16S539 in Dusun population of Eastern Malaysia. In this study, there was no detectable departure from the Hardy-Weinberg equilibrium observed for the three STRs loci by the Chi-square test. The data obtained in present study have been compared with STRs data of various population groups reported in the literature.

INTRODUCTION

Individual identification is a prerequisite in the salvation of violent crimes such as murder and rape, resolving unestablished paternity, identification of remains of missing persons, and victims of mass disasters (Butler, 2001). The technological development and advancement in various molecular biological applications have revolutionized concepts and knowledge in basic biology, medicine and forensic science. It has been 20 years since the first development of DNA fingerprinting and the start of forensic DNA typing (Tamaki and Jeffreys, 2005). DNA profiling technique becomes an important tool for human identification applications.

Prior to DNA fingerprinting, various methods used to attempt for positive identification of an individual. Alphonse Bertillon introduced a system known as "Bertillonage", also known as "anthropometry" in 1882, which used measurements of various parts of the body, including scars and other body marks. Johannes Purkinje first described the constancy of fingerprints in 1883, and the modern fingerprinting used in identification of individuals was introduced in 1892, by Sir Francis Galton. Karl Landsteiner published a report on the potential of blood groups to be used in forensic applications during 1903 (Wall, 2002). By using simple ABO blood groups, and the gradual discovery of new and different blood typing systems, the discriminatory power of blood groups in forensic applications has increased considerably. Though based on blood group polymorphism exclusion of individuals in a crime became certainty but not the inclusion (Lee and Gaensslen, 1990).

The DNA revolution in forensic investigation began in 1985 with the discovery of hypervariable minisatellite loci detectable with multi-locus probes (MLPs), by Alec Jeffreys and his co-worker (1985). They found that certain regions of DNA contained many dispersed tandem-repetitive 'minisatellite' regions, which are highly polymorphic

due to allelic variation in repeat copy number in the minisatellite. A technique was developed to examine the length variation of these DNA repeat sequence, by using a probe based on a tandem-repeat of the core sequence to detect many highly variable loci simultaneously (Jeffreys, 1985a). The results provide an individual-specific DNA 'fingerprint', as an important tool in human identification including parenthood testing. This DNA fingerprints are completely specific to an individual, except for his or her identical monozygotic twins (Jeffreys, 1985b). The tandem-repetitive regions known as variable number of tandem repeats (VNTRs). The technique used to examine the VNTRs in Jefferys' experiment was called restriction fragment length polymorphism (RFLP), which involved the use of a restriction enzyme to cut the regions of DNA surrounding the VNTRs (Samuels and Asplen, 2000; Wall, 2002; Aronson, 2005).

DNA fingerprinting helps to exonerate or convict suspects and to link a suspect's DNA to the DNA recovered at a crime scene. DNA fingerprinting was first used in a criminal case involving the sexual assault and murder of a schoolgirl in UK (Thieman and Palladino, 2004). Even though the case was solved successfully, but it highlighted one of the important limitations to the use of DNA evidence. The identity of the criminal cannot be established unless the known sample is available for comparison. DNA fingerprinting also helps in solving question of paternity. DNA analysis can theoretically match father to child with a much higher degree of accuracy, estimated to approach 99.9%. The first case solved in UK is the immigration tribunal hearing of a young British boy of Ghanaian descent (Aronson, 2005). DNA fingerprinting is also potentially exploited in forensic pathology and anthropology studies. The DNA profiling is usefully applied by the forensic pathologists in identification of exhumed skeleton. DNA profiling technique is useful in positively identification of remains during desert Storm, also the bodies and body parts of airline crash victims and other mass disaster. Biological sleuthing has become the new

wave of study in anthropology and evolution, as molecular biologists seek to derive and clone DNA from creatures including ancient human remains that are millions of years old (Allen and AuBuchon, 1995).

The various methods available for DNA analysis differ in their ability to differentiate two individuals and in the speed with which results can be obtained. The technologies used in DNA analysis can be broadly classified into technologies involving DNA probe and technologies involving the use of polymerase chain reaction (PCR) (Wall, 2002). The results all fall into the general description of DNA polymorphism, i.e. DNA that shows differences between individuals.

The earliest technique of direct DNA analysis available to forensic science is the measurement of variation in sequence length. This analysis is a multi-step procedure, and is broadly described as RFLP analysis. Depending on the type of probe used in the RFLP analysis, it will be described as multi-locus probe (MLP) analysis or single-locus probe (SLP) analysis (Wall, 2002; Jobling and Gill, 2004; Tamaki and Jeffreys, 2005). Human minisatellites or variable number tandem repeat (VNTR) loci have repeat units from 6 bp to more than 100 bp long depending on the locus, some show very high levels of allele length variability. These minisatellites were detected by hybridization of probes to Southern blots of restriction-enzyme-digested genomic DNA to reveal the RFLPs. MLPs used to detect many different minisatellites loci simultaneously, which shared a common 10 – 15 bp core GC-rich sequence, resulting in multi-band patterns, known as DNA fingerprints (Tamaki and Jeffreys, 2005). MLPs are highly variable between individuals, and used successfully in paternity testing and immigration cases, but it requires a great deal of labor, time and expertise to produce a good DNA fingerprints. Another disadvantage of MLP analysis is that it cannot be automated easily. On the other hand, MLP analysis requires several micrograms of high molecular weight genomic DNA,

whereas forensic specimens are often old and yield small quantities of degraded DNA. Therefore, MLP analysis is not suitable in all occasions in forensic work (Butler, 2001; Tamaki and Jeffreys, 2005). To overcome these limitations, specific cloned minisatellites were used as single-locus probes (SLP) to produce simpler DNA profiles. Since each SLP detects only a single minisatellite, it produces two bands (two alleles) patterns. Although the results are still highly polymorphic due to the use of hypervariable minisatellites, but the power of discrimination of SLPs are relatively low compares to MLP. Therefore, a panel of several SLP probes is normally used, with each producing a low probability of a match, collectively producing a relatively large probability when data is added together. The advantages of SLP analysis is far more sensitive, and mixed DNA samples such as semen in vaginal swabs can be analyzed (Tamaki and Jeffreys, 2005).

Some minisatellite loci with relatively short (~1000 bp) alleles can be amplified by PCR to yield amplified fragment length polymorphisms (AMPFLP). This method is used to measure the variation in a specific gene, with the presence or absence of arrays. D1S80 minisatellite locus was earlier used in forensic kinship testing which locus exhibits forty four alleles in populations (Tamaki and Jeffreys, 2005). The advantage of AMPFLP is the alleles amplified using PCR falls into discrete size and able to be categorized by the number of repeats. In contrast to RFLP analysis, which uses molecular markers to estimating the size of the DNA allele, a ladder containing a mixture of the possible alleles is run alongside the specimen DNA in AMPFLP analysis. The direct comparison can be done to yield the result easily (Allen and AuBuchon, 1995). The greatest disadvantage of this AMPFLP analysis is that they have a very limited range of alleles, or different types of DNA which can be measured. From the forensic point of view, the discriminatory power is too low.

Short tandem repeat (STR) profiling is the currently used DNA markers in human identification. STR analysis involves the variation in the copy numbers of the core repeats. STR s also known as microsatellites since the repeat unit is in the range of 2 – 7 base pair in length. STRs appear to be abundant throughout the human genome and occur at every 6-10 kb (Hochmeiser, 1995). Most microsatellite loci can be efficiently amplified by standard PCR since the repeat regions are shorter than 400 bp. The first forensic application of STR typing is from skeletal remains of a murder victim, followed by the identification of Josef Mengele, the Auschwitz "Angel of Death" (Tamaki and Jeffreys, 2005).

STR profiling is the state of the art technology used in human identification in most of the forensic laboratories. For this purpose the tetranucleotide or pentanucleotide STRs were validated taking into account factors such as constancy in repeat unit, robustness in amplification, and in low shutters. Further, the STRs used in forensic are amenable to multiplex PCR. Thereby in a single reaction many STRs can be amplified and analyzed. Furthermore, the fluorescent detection system have been developed in STR system make it amenable to automation of gel electrophoresis and DNA profile interpretation.

The use of single STR will only gives a very low discriminatory power. By combining several STR loci which are independent of each other, a high level of discrimination achieved. Currently 30 STRs loci were validated for forensic use (Hochmeiser, 1995). In the United Kingdom, 10 autosomal STR loci plus the amelogenin sex determination locus were used for compilation of database. The Combined DNA Index System (CODIS) was developed by Federal Bureau of Investigation (FBI) and the CODIS uses 13 STRs. The CODIS database consists of Convicted Offender Index, which contains profiles of individuals convicted of violent crimes, and the Forensic Index contains DNA

profiles from crime scene evidence, such as semen and blood. In 2003, the Japanese Police Agency introduced a 9 STR locus system, which uses some of the CODIS loci (the AmpFlSTRw Profilerw kit), for analysing forensic specimens (Samuels and Asplen, 2000; Tamaki and Jeffreys, 2005).

STR profiling is suitable for degraded DNA samples and even DNA recovered from a single cell, STR profiling can be achieved. STR systems are suitable for old and degraded DNA samples because of their small sizes. STRs are well characterized and have been used for typing of DNA extracted from cigarette butts, human remains and hair samples in forensic case work (Hochmeister *et al.,* 1995). STRs are highly discriminating between unrelated and even closely related individuals. Although STR system has been automated, for the laboratories which are not capable for purchase such expensive machine, for all multiplex systems, where the sizes of the three STR loci in the multiplex analysis do not overlap, can be typed manually by silver staining of denatured polyacrylamide gels. Because of the STR amplification products only differ by single repeat unit, it can be fully resolved by electrophoresis. The discrete alleles of STRs make results easier to interpret and to compare through the use of computerized DNA databases (Hochmeister *et al.,* 1995; Butler, 2001).

For application of STR profiling in criminal cases and for application in kinship testing, it is essential to compile STR DNA database for each of the population group in a country. There is no population database of STR III of Dusun population in Sabah, Eastern Malaysia has been studied and published. Hence the objective of present study is concentrated to study the allelic distribution and therefore compile a database of STRs loci D7S820, D13S317 and D16S539 of Dusun population.

REVIEW OF LITERATURE

The study of genetics started 1858, when Charles R. Darwin and Alfred R. Wallace published the theories of evolution by natural selection. In 1865, Gregor Mendel discovered the laws of heredity through the breeding experiments of peas. He contributed a mathematical model by which inherited characteristics are transmitted from parents to offspring. This theory later is termed as Mendel's Law, also known as theory of transmissible factors (Alcamo, 1996; Lewis, 2003; Dahm, 2005). The foundation of cytogenetics laid in 1866, when Haeckel proposed that the nucleus contained the factors responsible for the transmission of hereditary traits (Moore, 1972). Later, Flemming described the morphology and behavior of the chromosomes during cell division, coining both the terms "chromatin" and "mitosis" (Biotechnology Institute, 2005). In 1902, W. H. Sutton pointed out that chromosomes and inheritance factors which Mendel proposed responsible for heredity could be identical. Further, in 1903, Sutton suggested that only a part of chromosome is the basis for a trait, whereas the chromosome may be divisible into smaller entities. Thomas Hunt Morgan (1910) demonstrated a single trait related to a single chromosome in the functioning of the cell based on his studies on the fruit fly *Drosophila melanogaster* in his experiment (Alcamo, 1996).

The history of DNA started with the discovery of Johann Friedrich Miescher, who isolated the hereditary material in 1869 (Dahm, 2005). He separated nucleus from human white cell and searched for protein. He discovered a new substance which chemically unlike any known class of chemicals, and termed it as nuclein (Alcamo, 1996). In 1879, Albrecht Kossel discovered that nuclein was comprised of four bases and sugar molecules. The inheritance factors were termed as gene by Willard Johannsen (1910). 1920s, Phoebus A. T. Levene determined the basic chemistry of nuclein, and coined the name deoxyribonucleic acid (DNA) (Alcamo, 1996). In 1928, Frederick Griffith performed

experiments with *Streptococcus pneumoniae* and postulated some substances in S strain bacteria were entering the R strain and transforming them, which called "transforming principle" (Benjamin, 1997). James Lionel Alloway and his co-worker demonstrated the extraction of the transforming principle from the mixture with alcohol (Alcamo, 1996; Benjamin, 1997).

In 1945, Oswald T. Avery and his associates Colin MacLeod and Maclyn McCart had established the chemical nature of the transforming principle, DNA. Hershey and Chase (1952) performed experiments with bacteria and viruses that replicate within them, and provided the final proof for DNA's involvement in heredity (Alcamo, 1996). Watson and Crick completely described the double helix structure of DNA, with its complementary base-pairing, in 1953 (Watson and Crick, 1953). During 1961, Robert W. Holley and his co-worker deciphered the genetic codes (Lewis, 2003; Klug, 2004; Dahm, 2005). In 1972, a breakthrough occurred with the development of the recombinant DNA technology, when Paul Berg created the first piece of recombinant DNA used restriction enzymes. Frederick Sanger and his collages first developed the method to sequence DNA in 1977. Six years later, Kary Mullis invented PCR as a method for amplifying DNA in vitro (Freifelder and Malacinski, 2002; Dahn, 2005).

The evolution of forensic genetics began more than a century ago. The ABO blood groups polymorphism, was discovered by Karl Landsteiner in 1900, and found applicable to solving crimes. Until 1970s, 17 blood group systems and few serum proteins and enzymes validated but not all useful in forensic works (Samuels and Asplen, 2000). The main reason is these markers were with low variation and were degraded rapidly by bacterial enzymes. The match probability obtained was 0.01 – 0.001. Further, all markers are not present in other body fluids. This marker system also not applicable on mixed samples encountered in rape cases (Jobling and Gill, 2004).

In criminal investigation, a higher standard of proof is required. By discovery of DNA fingerprinting by Alec Jeffrey (1985), this requirement had been fulfilled. Alec Jeffrey first demonstrated the use of DNA in criminal investigation (Aronson, 2005; Samuels and Asplen, 2000). The individual variability in the DNA is much greater compared to serological and enzymatic markers. With using only a single probe in DNA fingerprinting, the match probability was estimated to be 3×10^{-11}, when two probes used together the match probability increased to 5×10^{-19} (Jobling and Gill, 2004).

The advantages of DNA methods in forensic works included the tests are based directly on the genetic makeup, the DNA, of the individual. Therefore, the complication from dominance and recessiveness, which is generally faced when using the serological and protein tests to identify the gene product, is avoided (Samuels and Asplen, 2000). Furthermore, since DNA is found in cells throughout the body, all biological materials collected at the crime scene can be analyzed. The great advantage of DNA markers is that, it has much greater stability than blood group markers used in forensic application previously (Allen and AuBuchon, 1995; Wall, 2000; Samuels and Asplen, 2000).

The DNA is contained in 23 pairs of chromosomes present in human diploid cell. Out of the 23 pairs of chromosomes 22 pairs are autosomes and the remaining pair XX in the female or XY in the male is the sex chromosome. The DNA of two individuals except for identical twins – monozygotic twins differ considerably due to meiotic reshuffling of DNA (Lewis, 2003).

The 46 chromosomes in each human cell consists two complete sets of genetic information. The human genome probably contains from 28,000 to 34,000 protein-encoding genes, scattered among three billion DNA bases among each set of 23 chromosomes (Lewis, 2003). Thus, a body cell has approximately 6 billion bases. The precise processes of DNA duplication and cell division ensure that each cell contains the

same sequence of DNA bases. DNA samples from two unrelated persons differ on the average at only about one base per thousand. Yet 1/1,000 of 6 billion is 6 million, which are sufficient to produce the genetic differences among two persons. In addition, there are also some variations in the individual nucleotides number, due to the variable number of times of some DNA regions in which a small number of bases is repeated. Therefore, the total amount of DNA in different individuals is not exactly the same. Forensic scientists shows great interest in the regions where the repeated sequences in which the number of repeats varies from person to person (Samuels and Asplen, 2000).

Watson and Crick analyzed the x-ray diffraction of DNA molecules and figured out that DNA is a double-stranded twisted helical structure (Watson and Crick, 1953). The bases of one strand are hydrogen-bonded to another strand to form the purine-to-pyrimidine base pairs of A-T and G-C. They are stacked with their planes separated by a spacing of 3.4 Å. The ordering of the nucleotides defines DNA sequence, thus the protein they encode (Butler, 2001; Freifelder and Malacinski, 2002). A stretch of DNA from less than hundred to more than million base pair in length called gene. A gene responsible for expression for a particular characteristic, for example encode for a protein production. The variants of gene are called alleles (Lewis, 2003). An individual with two identical alleles of a gene is called homozygous, and with two different alleles is called heterozygous. Alleles which only express their characteristic trait when present in duplicate are recessive. Dominant alleles are those which effective when present in single copy. Genotype is the genetic makeup of an individual. It may be included several gene loci. Phenotype is the trait, which could be observed externally or measurable. The genotype of the homozygous individual directly reflected by phenotype, but for heterozygous individual, it depends on the relationship between the types of alleles present (Benjamin, 1997). Locus (plural is loci) is the term describes the specific position of a gene at a specific chromosome. There

are more than two different alleles at a locus in a population. A locus with more than one allele in the population is said to be polymorphic and a variant in sequence that is present in at least one percent of a population is called a polymorphism. The highly polymorphic loci are very useful in forensic individual identification (Samuels and Asplen, 2000; Lewis, 2003).

Genes are transmitted from generation to generation unchanged. Somehow, the changes in gene sequence could occur due to mutation. Mutation refers to any change in the base sequence of DNA, included substitution, addition, rearrangement, or deletion of one or more bases. It may also involve chromosome, where sometimes broken and reattached in new way, or the whole chromosome may lost or duplicated. The mutagenesis could be spontaneous or induced. Once mutates, the mutant form of gene is stable, and transmitted to the offspring (Freifelder and Malacinski, 2002). This mechanism serves for more polymorphic loci in a population.

Human genome may divide into many different classes based on the functional properties. The coding region contained genes which carry the information for protein products expressed in different cells and tissues. About 97% of human genome made up of non-coding region, contains DNA sequences which have no function, and these are referred as "junk DNA" (Fowler, 1989; Samuels and Asplen, 2000). The non-coding region shows the same genetic variability as genes do, somehow, could be more. The markers used for forensic identification usually located at this non-coding region, or in the non-coding part of a gene named intron. The genotype for the group of loci used in the forensic DNA analysis is called DNA profile (Samuels and Asplen, 2000).

The DNA sequences in non-coding region may exist in single copy or multiple copies, which are termed repetitive DNA. These repetitive sequences contribute for 20 to 30 per cent of human genome, broadly classified as tandemly repetitive sequences and

interspersed repetitive sequences. Tandemly repetitive sequences are now further classified into macrosatellite, minisatellite and microsatellite DNA (Fowler, 1989). Macrosatellite DNA contains repeat units of up to 200 bp arranged in clusters between 100 and 5000 kbp, predominantly distributed at centromeric and pericentromeric DNA. Minisatellite has medium length repeat, with the core repeat unit in range of 10 to 100 bp in length. Microsatellite is the DNA regions with repeat units that are 2 – 6 bp in length. Dispersed between the centre and tips of human chromosomes are interspersed tandem repeats, as well as short interspersed elements (SINEs, <500 bp), and long interspersed elements (LINEs, > 500 bp). These repetitive sequences typically repeated between a few and several thousand times, resulting in multiplicity of lengths, called length polymorphism (Fowler, 1989; Hochmeister, 1995; Butler, 2001).

Alec Jeffreys (1985) used the technique known as restriction fragment length polymorphism (RFLP) to detect the variation in repeat copy number of dispersed tandem-repetitive minisatellite regions, later termed variable number of tandem repeats (VNTRs) (Lewis, 2003; Tamaki and Jeffreys, 2005). Although RFLP based DNA fingerprinting produced strong evidence for individual identification and paternity testing, but problems arise due to the often degraded DNA samples in forensic cases. RFLP technique need at least 5 – 10 μg nondegraded DNA (Benecke, 1997). Edward *et. al.* (1991) has introduced short tandem repeats (STRs) based DNA profiling using polymerase chain reaction (PCR), enabling DNA profiling of degraded DNA samples. STRs spread over the entire genome and are situated exclusively in non-coding region. Every person's genome estimated contains up to hundreds of STR loci, and for every locus, there are generally 5 – 10 possible alleles observed (Benecke, 1997), sometimes may up to 30 alleles (Samuels and Asplen, 2000). By the comparison of the alleles of a person (STR profile) to another, individuality can be established (Benecke, 1997). The selection criteria of STR loci in

human identification include the high discrimination power, usually more than 90%; located in separate chromosome to avoid problems of linkage between the markers; robustness and reproducibility; low mutation rate; and with discrete length of alleles that fall in range of 90 – 500 bp (Gill *et. al*, 1996).

A common nomenclature for STR locus developed to aid in inter-laboratory reproducibility and comparison of data in the forensic DNA community. The Commission of the International Society of Forensic Haemogenetics (ISFH) had issued the guidelines for designating STR alleles in 1994 and 1997 (Butler, 2001). Allelic ladders containing sequenced alleles that are named according to the ISFH recommendation should be used as the reference for allele designation in unknown samples. An allelic ladder is an artificial mixture of the common alleles present in the human population for a particular STR marker. Allelic ladders served as a standard for measuring the stick for each STR locus.

To contribute a STR profile, the STR loci are amplified using PCR technique. Nowadays, at least 16 STRs can be amplified in a single reaction (Butler, 2001). The amplified products are separated by electrophoresis. There are two detection system used in forensic application, silver staining detection and laser detection. In silver staining method, the entire gel is stained with silver, allowed silver to bind to DNA. The DNA bands become visible after developed in developing solution. The second method used the primers tagged with fluorescent, allowed it to incorporate into the STR fragments generated during the amplification step. After the separation of fragments by electrophoresis, instrument is used to detect the position of separated fluoresce products. This method become increasing prevalent since the automation facilitated the STR profiling procedure (Benecke, 1997; Samuels and Asplen, 2000).

The use of DNA evidence and convicted offender DNA databases has expanded significantly. Population databases proved particularly useful for preventing and solving

crime. This DNA databases used are designed to be representative of the entire population. The variation in allele frequency across population subgroups can be characterized using statistical methods developed by population geneticists. Because these subgroups are ill-defined, geographic sampling method is adopted. It has been found that any particular allele can have a frequency that differs significantly from one population to another. For maximum efficiency in criminal investigation, the establishment of population databases compiled a worldwide survey derived from databases compiled by forensic scientists around the world is demanded. Therefore, the comparison of the allele frequencies in different recognizable geographic groups is made possible (Reynolds, 1995; Benecke, 1997).

OBJECTIVE OF THE STUDY

For application of STR profiling in criminal cases and for application in kinship testing, it is essential to compile STR DNA database for each of the population group in a country. There is no population database of STR III of Dusun population in Sabah, Eastern Malaysia has been studied and published. Hence the objectives of present study are:

i. To study the allelic distribution of STRs loci D7S820, D13S317 and D16S539 of Dusun population.

ii. To compile a database of STRs loci studied of Dusun population for application in forensic caseworks.

MATERIALS AND METHODS

MATERIALS

All materials and reagents used are sterile. The STR typing procedure was done as the guideline of manufacturer of STR kits used (*GenePrint*® STR Systems (Silver Stain Detection) Technical Manual D004, Promega, USA).

Reagents

1 M Tris-HCl pH 7.5, 0.5 M EDTA pH 8.0, 20% SDS, digestion buffer, proteinase K (Promega, USA), buffered Phenol (PIERCE, USA), Chloroform:Isoamyl alcohol (24:1), chilled ethanol, 2 M sodium acetate, 70% ethanol, TE buffer, 1 X TBE buffer, 5 X loading solution, Ethidium bromide, STR 10 X Buffer (500 mM KCl, 100 mM Tris-HCl (pH 9.0) at 25°C), 10 X STR III Primer Pair Mix, *Taq* DNA Polymerase (5u/µL), K562 DNA (2µg/µL), 95% ethanol, Rain-X repellant, bind saline (Promega, USA), 10 X TBE buffer, 40% acrylamide:bis (19:1), 10% ammonium persulphate, 0.5 X TBE, STR 2 X loading Solution (10 mM NaOH, 95% formamide, 0.05% bromophenol blue, 0.05% xylene cyanol FF), STR allelic ladder (Promega, USA), fix/stop solution, silver nitrate staining solution, developer solution.

Chemicals

Tris base (Promega, USA), concentrated HCl, Na_2EDTA (Promega, USA), NaOH (Merck, Germany), Sodium dodecyl sulphate (Bio-Rad Lab), chloroform (Merck, Germany), Isoamyl alcohol (Merck, Germany), sodium acetate (Merck, Germany), glacial acetic acid (Merck, Germany), absolute ethanol (Merck, Germany), Tris-HCl (Promega, USA), EDTA (Promega, USA), boric acid (Promega, USA), ethidium bromide (Bio-Rad Lab), agarose (Promega, USA), acrylamide (Promega, USA), bisacrylamide (Promega, USA),

ammonium persulfate (Promega, USA), TEMED (Promega, USA), urea (Promega, USA), silver nitrate (Promega, USA), Formaldehide (Promega, USA), sodium-thiosulfate (Promega, USA), sodium carbonate (Promega, USA).

Apparatus

Autoclave, microcentrifuge, water bath, balance, magnetic stirrer, Gilson varying volume pipette (1000 µL, 200 µL, 100 µL, 20 µL and 10 µL), vortex, parafilm, vacuum pump, microwave oven, freezer, refrigerator, agarose gel apparatus (Bio-Rad Lab), high voltage power supply, UV light illuminator, MJ Research PTC-200 Peltier Thermal Cycler, SA 32 Electrophoresis apparatus (GIBCO BRL Sequencing System), plastic trays.

METHODS

Buccal Swab Collection

Buccal swabs were taken from 100 healthy unrelated random Dusun individuals within Malaysia by using sterile cotton buds. Formal consent was obtained from each individual. Two sticks of buccal swab samples were collected from each individual by rubbing the cotton bud against the inner side of the individual's cheek. The cotton buds were air dried and kept in the plastic envelopes, labeled with details of collection. The samples were kept in dry environment at room temperature.

Buccal Swab DNA Extraction

Reagents Preparation

Proteinase K

Proteinase K solution was prepared dissolved 50 mg of Proteinase K in 5 mL of sterile deionized water.

1 M Tris-HCl pH 7.5

Tris-HCl (1 M) (pH 7.5) solution was prepared by dissolved 121.1 g of Tris-base in 800 mL of deionized water. The pH was adjusted to 7.5 by adding few drops of concentrated HCl. The solution was made up to 1000 mL with deionized water and autoclaved.

0.5 M EDTA pH 8.0

EDTA (0.5 M) (pH 8.0) solution was prepared by dissolved 186.1 g of Na_2EDTA in 800 mL of deionized water. The solution was stirred vigorously with a magnetic stirrer. The pH was adjusted to 8.0 with NaOH pellet. The solution was made up to 1000 mL with deionized water and autoclaved.

10% SDS

SDS (10%) solution was prepared by dissolved 10 g of Sodium dodecyl sulphate in 80 mL of deionized water. The solution was stirred with heat using magnetic stirrer. Few drops of concentrated HCl were added to adjust to pH 7.2, and made up to 100 mL with deionized water and stored in an autoclaved bottle.

Digestion buffer

Digestion buffer was prepared by added 1 mL of 1 M Tris-HCl pH 7.5, 2 mL of 0.5 M EDTA, 20 mL of 10% SDS, and 1 mL of NaCl to 60 mL of deionized water. The solution was mixed well and made up to 100 mL with deionized water.

Chloroform:Isoamyl alcohol

Chloroform:Isoamyl alcohol solution was prepared by added 10 mL of Isoamyl alcohol to 240 mL of chloroform and mixed well.

3 M sodium acetate

Sodium acetate (3 M) solution was prepared by dissolved 102.025 g of sodium acetate in 200 mL of deionized water. The pH was adjusted to 7.2 with glacial acetic acid. The solution was made up to 250 mL and autoclaved.

2 M sodium acetate

Sodium acetate (2 M) solution was prepared by added 16 mL of 3 M sodium acetate to 8 mL of deionized water.

70% ethanol

Ethanol (70%) solution was prepared by added 350 mL of absolute ethanol to 150 mL of deionized water.

TE buffer

TE buffer was prepared by added 10 mL of 1 M Tris-HCl and 0.2 mL of 0.5 M EDTA to 989.8 mL of deionized water and autoclaved.

Extraction Method

The dried cotton buds were cut and placed in a 1.5 mL microfuge tube. 500 µL of digestion buffer and 12 µL of Proteinase K (20 mg/mL) were added into each tube. The samples were incubated at 56°C overnight. 120 µL buffered phenol was added and mixed vigorously using vortex. The samples were centrifuged at 10,000 rpm for 3 minutes. The supernatant of the sample was transferred to another 1.5 mL microfuge tube by using the cut tip. One volume (250 µL) of buffered phenol and chloroform-isoamyl was added to the supernatant. It was mixed vigorously and centrifuged at 10,000 rpm for 3 minutes. The