UNIVERSITI SAINS MALAYSIA

UNIVERSITI SAINS MALAYSIA

Polymorphism of CSF1PO, TH01 and TPOX in random Bajau Population of Malaysia

Dissertation submitted in partial fulfillment for the

Degree of Bachelor of Science (Health) in Forensic Science

Lelyan Binti Minsun

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 CSF1PO, THO1 and TPOX in random Bajau Population of Malaysia"

Is the bona fide record of research work done by

Ms. LELYAN BINTI MINSUN

During the period 26 December 2005 to 30 Mac 2006

Under my/our supervision

Signature of Supervisor: .....

Name and address of Supervisor: Mr. S. Panneerchelvam,
Lecturer,
School of Health Sciences,
Universiti Sains Malaysia,
16150, Kubang Kerian, Kelantan.

Date: 11/05/06

Signature of Co-Supervisor: .....

Name and address of Co-Supervisor: Prof. Dr. Norazmi Mohd Nor,
Lecturer,
School of Health Sciences,
Universiti Sains Malaysia,
16150, Kubang Kerian, Kelantan.

Date:

ACKNOWLEDGEMENTS

In preparation of this thesis paper, I would like to express my appreciation to all lecturers, friends and my parents who have supported and responded my work either directly in discussions of the material or indirectly by their encouragement and patience.

I wish to acknowledge the kind support of the project supervisor Mr. S. Panneerchelvam who kindly gave me an idea, support and opinion in doing the research project.

Also lot of appreciation to Puan Nor Haslinda as her concern to guide me for completing my laboratory work and all the skill she has been given, really valuable to me as I can completing my research earlier. Not to forget, special thanks to Puan Rosniah and Mr. Aiman Asyraf for their help and assistance in completing my paperwork.

Thanks also to all my responder who are willingly to give their cooperation in giving their saliva samples to me. Your good deeds will not be ever forgotten. Also warmest thanks are in order to my friends and parents, respectively, who were exceedingly patient with me while I was completing the paper as well as providing help and encouragement in this effort.

Last but not least, to those who have contributed directly or indirect but their names were not mentioned here thank you to you all. Hope the present study will give benefit to people who encounter it.

Thanks again.

CONTENTS

LIST OF TABLES AND FIGURES

List of Tables

List of figures

ABSTRACT

Genotype polymorphism studies at three short tandems repeat (STR) loci (CSF1PO, TPOX, TH01) were carried out in one ethnic (Bajau) living in Sabah, Malaysia. The analysis was performed on 150 unrelated healthy individuals belonging to Bajau races. The allele frequency of each locus was calculated from the observed number of each genotype. All loci were in accordance with Hardy-Weinberg equilibrium (P = 0.05). No deviations from equilibrium were observed. The discriminatory power and exclusion probability values for all the analyzed markers are significantly high and thus reveal high forensic significance. Statistical analysis was also carried out to obtain some parameters of medico-legal interest and comparative studies were carried out with other populations studied to date for these loci. This allele frequency data will be useful for forensic analyses and paternity analysis to estimate the frequency of a multiple STR locus DNA profile in Bajau population.

INTRODUCTION

Deoxyribonucleic acid (DNA) constitutes the primary genetic material found in the cells of all living things and occurs predominantly in the nucleus of cells. The human genome consists of about three billion base pairs along 46 DNA molecules contained in 23 pairs of chromosomes. Approximately 5% of human genome is a coding sequence, which is dedicated to our species-specific traits. The remaining 95% of the human genome is non-coding sequences, the individual unique that provides the basis for distinguishing between individuals through DNA profiling (Jain, 1999). DNA profiling, also called DNA fingerprinting or DNA typing is the process of sorting out an individual's unique, polymorphic, fragments from the common ones (Krawczak and Schmidtke, 1998). In other words, it is the biological tool in the analysis of DNA that allows scientists to compare samples of DNA material (Rudin and Inman, 2002).

Sources of DNA material for testing may be blood, hair roots, semen stains, tissues, saliva, urine or any body fluid in human body. In 1980, American geneticists discovered DNA sequences that do not hold any genetic information and which is extremely variable between individuals. Variations in inherited regions of DNA sequence between individuals are termed "polymorphism" (Brown, 2001). The DNA polymorphism can serve as markers on the genetic chromosome and consists of several types; restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), microsatellite repeats or variable number of tandem repeats (VNTR), single strand conformation polymorphism (SSCP) and single nucleotide polymorphism (SNP). All these DNA polymorphisms are very useful

markers for DNA profiling, human genetics, biomedical and forensic analysis (Dale and Schantz, 2002). The sequences with the highest degree of polymorphism are very useful for DNA analysis in forensic cases such as criminal case and paternity testing (Weaver, 2005).

In criminal cases, the use of DNA profiling is based on the fact that it is extremely unlikely that any two unrelated individuals own identical DNA. Investigators may compare DNA found at the crime scene with DNA taken from a victim involved or suspected criminal or compare to a database of DNA profiling results from other scenes or from convicted offenders. Investigators use DNA profiling as supported evidence to link to, or exclude from, that scene, a possible suspect. If there is a 'match', the two samples may have comes from the same person. But it is only be useful if there is enough DNA in the samples. In paternity cases, the biological father of a child is determined on the fact that a child's DNA has some of the heritable DNA profiling characteristics from that of the father (Rudin and Inman, 2002).

Each person has a unique pattern of DNA profiling with the exception of identical siblings that makes it so valuable evidence. Sir Alec Jeffreys and his colleagues (1985) were the first group to perform DNA profiling using RFLP (Weaver, 2005). They discovered the regions of the genome in a specific sequence of base pairs is repeated a variable number of times in a continuous sequence. These regions are called VNTR regions. The length of a given VNTR region of a certain chromosome will thus depend on the number of base pairs in the repeat sequence and the number of times it is repeated. In DNA profiling, the lengths of a number of different VNTR regions are determined and the DNA profile is thus a simple combination of these lengths.

A RFLP is a sequence of DNA that has a restriction site on each end with a 'target' sequence in between. A target sequence is any segment of DNA that binds to a probe by forming complementary base pairs. A probe is a sequence of single stranded DNA that has been tagged with radioactivity or an enzyme so that the probe can be detected. When a probe bind to its target, we can detect this binding and know where the target sequence is since the probe is detectable. RFLP produces a series of bands when a Southern blot is performed with a particular combination of restriction enzymes and probe sequences. When using Southern blots to detect RFLPs, it will mainly identify larger-scale changes on the structure of the genome, or a specific part of it, and much less likely to identify minor sequence changes in a specific gene. So, RFLP method has limitations because it requires large amount of DNA for the signal to be detectable. This approach is not possible for limited amount of DNA and has to use PCR to amplify the tiny amount of DNA that is available (Dale and Schantz, 2002).

Single Nucleotide Polymorphism (SNP) is the simplest form of genetic variation consists of a change in the sequence of bases at the single point. These point mutations are typed by analysis with short oligonucleotides probes that hybridize to the alternative forms of the SNP. The number of SNPs in the human genome is not yet known but is at least 1.4 million (Brown, 2001). Mitochondrial analysis are based on mitochondrial DNA that contains polymorphisms which can be used to infer relationships between individuals, but the degree of variability is not as great as displayed by autosomal STRs. So mitochondrial DNA is not routinely used for the analysis of forensic specimen. But mitochondrial DNA has the important property of being inherited solely through the female line, the father's mitochondrial DNA being lost during fertilization and not contributing to the

son or daughter's DNA content. This maternal inheritance pattern makes it easier to distinguish relationships when the individuals being compared are more distantly related (Brown, 2001).

DNA analysis can also be used to identify the sex of an individual. The genetic difference between the sexes is the possession of a Y chromosome by male, so detection of DNA specific for the Y chromosome would enable males and females to be distinguished. Analysis in determining sex is deal with bodies that are so badly damaged. In the analysis of archaeological specimen this technique is used for determining the sex of the human remains. In particular, there are several repeated sequences that are only located in the Y chromosome, these repeated sequences acting as multiple targets for the PCR and hence giving greater sensitivity, an important consideration when dealing with a badly damaged body or an ancient bone. A PCR directed at Y-specific DNA sequences would give a product with male DNA but no band if the sample comes from a female (Brown, 2001).

The other types of DNA profiling known as modern forensic DNA profiling uses short tandem repeats (STR) analysis. The repeat unit for STRs can vary in between two to seven base pairs long. DNA sequences for STRs region are amplified by using the polymerase chain reaction (PCR) technique, which is more sensitive and efficient method. Through this technique, it is possible to produce several million copies of a specific DNA sequence within a few hours. The amount of DNA samples required for STR DNA profiling is small and even degraded samples can be efficiently profiled. STRs genetic markers are widely used in forensics, gene mapping, medicine and anthropological studies (Dale and Schantz, 2002).

The number of repeats in a particular STR is variable because repeats can be added or, less frequently, removed by errors that occur during DNA replication. In the human genome, the most common type of STR is the dinucleotide repeat $[CA]_n$, where 'n', the number of repeats, is between 5 and 20. In the population as a whole, there might be as many as ten different versions of a particular STR, each of the alleles characterized by a different number of repeats. Importantly, DNA profiling when directed at STRs with large numbers of alleles, gives a very high statistical probability that a match between a test profile and that of a suspect is significant and not due to a chance similarity between two different people. The necessary degree of certainty can be achieved by analysis of a panel of nine STRs, which can be typed in a single multiplex PCR, in which a series of primers pairs are used in a single reaction (Brown, 2001).

Modern DNA profiling is both faster and less labor intensive as many of the steps were automated. It is possible in a single reaction to amplify DNA from fifteen to twenty STR regions. The PCR technique also makes it achievable to carry out analyses on samples containing very little DNA. Classical DNA profiling needs at least micrograms of high molecular weight DNA, while at least one nanogram is adequate for a routine PCR-based analysis, although a single cell (less than 10 picograms) can be enough (Brown, 2001).

The uniqueness of human DNA especially the microsatellites VNTRs regions is used to create databases of genetic profiles of convicted criminals by the law enforcement agencies. There are several types of databases that have been studied such as population DNA databases, criminal DNA databases, human sequence databases, and other databases such as protein databases and nucleotide/sequence databases. STR database for Malaysian Malays, Chinese

and Indian were already available (Seah *et al.*, 2003; Panneerchelvam *et al.*, 2003; 2004).

This particular study aimed to compile STR DNA database for Bajau population of Eastern Malaysia. The Bajau ethnic population of Sabah has not been studied yet for the purpose of population DNA database. The main objective in this research project is to study the pattern and distribution of three validated STRs loci in Bajau population of Malaysia. All these loci are validated and routinely used in similar studies throughout the world.

REVIEW OF LITERATURE

Cell is the fundamental, structural and functional unit of living organisms. All living things are made up of cells. The human body has more than 10 trillion cells. A cell is the smallest unit of life that can exist independently and seen only with a microscopy. Robert Hooke first observed the cell under microscope in 1665. He saw that it was composed of neat holes enclosed by walls and called these holes cells. During 1838, Matthias Schleiden stated that the cell was the basic unit of all life. Theodor Schwann (1839) a zoologist formulated the theory that all living things are made up of cell. Both of them proposed the cell theory that includes of three basic principles; all living things are made up of one or more cells, the smallest living unit of structure and function of all organisms is the cell and all cells arise from preexisting cells. And from that time biologist regarded the cell as the building block of life (Alters, 2000).

Gregor Mendel formulated the basic laws of heredity through experiments with garden peas. He suggested that traits are inherited as discrete packets of information. He studied traits in pea plants. These studies led Mendel to formulate the first correct theory of heredity. His theory had two principles called Mendel's laws of heredity. Mendel's first law, the Law of Segregation, has three parts; hereditary characteristics are determined by separate units (now called genes), these units occur in pairs and the genes in a pair segregate (separate) during the division of sex cells, and each sperm or egg receives only one member of the pair. Mendel's second law is called the Law of Independent Assortment. It states that each pair of genes behaves independently of all other pairs in the production of sex cells. Therefore, each gene pair is inherited independently of all other genes.

Geneticists now know that independent assortment applies only to genes that are on different chromosomes or far apart on the same chromosome. Genes that are linked, or near each other on the same chromosome, tend to be inherited together (Klug and Cummings, 2000).

In 1900's, Hugo de Vries of the Netherlands, Carl Correns of Germany and Erich von Tschermak of Austria were discovered Mendel's work by working independently on the problem of heredity. In 1902, Walter S. Sutton pointed out that during cell division chromosomes behaved, as Mendel had believed inherited traits behaved. In 1910, Thomas Hunt Morgan and his co-workers proved that genes are the units of heredity (Alters, 2000). They also proved that genes are arranged in a specific order in chromosomes. Morgan studied the laws of heredity by using the fruit fly (*Drosophila melanogaster*) for experiments in breeding. His research clarified the physical basis for the linkage and recombination of heredity traits. He was the first to explain sex-linked inheritance that some traits pass to only one or the other sex. Morgan and his associates proved that genes are arranged on the chromosomes in a specific order (Alters, 2000).

Beadle and Tatum (Klug and Cummings, 2000) discovered that some genes control chemical reactions in cells by directing the formation of enzymes. They found that there is a specific gene functions to produce one enzyme and developed the one-gene-one enzyme theory (Klug and Cummings, 2000). In 1940's, scientists are more interested in the chemistry of genes. They knew that chromosomes consisted of DNA and protein. In 1868, Friedrich Miescher discovered in the cell nucleus a mixture of compounds that he called nuclein. The major component of nuclein is deoxyribonucleic acid (DNA). But scientists had dismissed DNA as unimportant, knowing how essential proteins were in life

processes. The turning point came in 1944; when a team headed by Oswald T. Avery found evidence that DNA alone determine heredity (Weaver, 2005).

In 1953, Watson and Crick (Alters, 2000) proposed that the structure of the DNA molecule resembled a twisted ladder. They based this model on the experimental findings of Rosalind E. Franklin and Maurice H. F. Wilkins. Watson and Crick proposed that DNA is a double helix – two DNA strands wound around each other. The bases of each strand are on the inside of the helix, and a base on one strand pairs with one on the other in a very specific way. DNA has only four different bases: adenine (A), guanine (G), cytosine (C), and thymine (T). This complementarity's is what allows DNA to replicate faithfully. The two strands come apart, and enzymes build new partners for them using the old strands as templates and following the Watson-Crick base pairing rules (A with T, G with C). This is called semi conservative replication because one strand of the parental double helix is conserved in each of the daughter double helices. Experiments have proved their model correct (Alters, 2000).

Many scientists have worked on unraveling the genetic code, found in the sequence of the bases of DNA. Marshall W. Nirenberg (1962) discovered the code for one amino acid-UUU, which served as a cornerstone for the complete analysis of the genetic code. By 1967, he and others eventually determined the code for the 20 amino acids involved in protein production (Klug and Cummings, 2000). Genetic engineering techniques developed in mid-1970, offers many potential benefits in medicine, industry and agriculture especially might help scientists to learn more about the structure and function of genes. These techniques are also called recombinant DNA technology. Recombinant DNA

technology alters genes in organisms to make molecules called recombinant DNA (Alters, 2000).

Through development of recombinant DNA technology offered tools for forensic identification. The forensic analysis of DNA is perhaps the most significant advance in modern crime solving and law enforcement. In 1980, David Botstein and co-workers were the first to exploit the small variations found between people at the genetic level as landmarks to construct a human gene map. The particular type of variation they used is called restriction fragment length polymorphism (RFLP). RFLPs are inherited variations in the size of DNA fragments produced when DNA is digested with restriction endonucleases and provide a considerable degree of individual variation, which is inherited according to Mendelian laws (Lee & Gaensslen, 1990).

Alec Jeffreys (1984), while searching for disease markers in DNA discovered a unique application of RFLP technology to the science of personal identification. His method was modified to detect loci sequentially rather than concomitantly, and was adopted by crime laboratories for general use in the United States. Kary Mullis (1986) discovered polymerase chain reaction (PCR) method of replicating particular regions of a DNA molecule. One of the major advantages of PCR is that it enables amplification (from an extremely small quantity of DNA) of a specific sequence, which does not even have to be pure form. PCR, more than any other scientific advance since perhaps the elucidation of the structure of DNA, has changed the face of molecular biology. Most crime laboratories today employ PCR-based short tandem repeat repeat DNA typing (Rudin & Inman, 2002).

The advance of DNA profiling through PCR-based STR can be used in forensic testing during investigation of answering question whether or not a given individual under suspicion can be matched to material found at the site of the crime. In cases of a match, however, when the degree of resemblance between trace and suspect DNA is compatible with identical origins, the question remains whether the observed coincidence could also be due to chance. Thus, estimates of genotype frequencies are required, and statistical analyses are involved. The weight which can be attributed to a match between two DNA profiles depends on how likely it is for a person, picked at random from the population, to show the same profile. A large number of individuals have therefore been analyzed in various countries in order to acquire information about allele frequencies (Krawczak and Schmidtke, 1998).

OBJECTIVE OF THE STUDY

- To develop a database for three STRs loci – CSF1PO, TPOX, TH01 in Bajau population of Eastern Malaysia.
- To compare the allelic distribution pattern of CSF1PO, TPOX and TH01 for Bajau population with other populations in the world.

MATERIALS AND METHODS

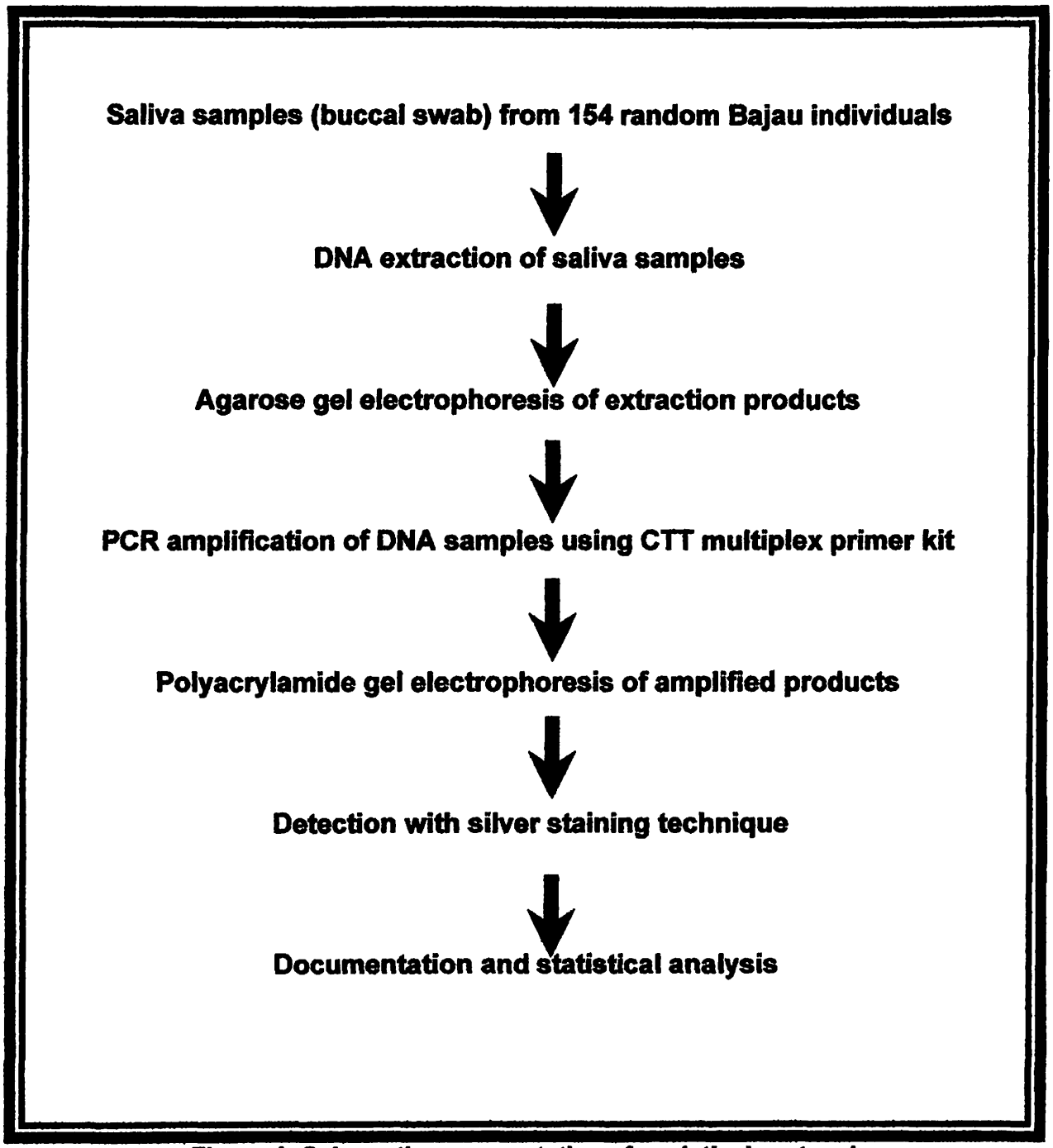

Figure 1: Schematic representation of analytical protocol

MATERIALS

1. Sample Source

Polymorphism of CSF1PO, TH01 and TPOX STR loci in Bajau population of Sabah, Malaysia was determined by saliva swab samples from 150 unrelated volunteers. The subject was requested to rub their inner cheek with cotton buds. The cotton buds (buccal swab sample) were dried at room temperature and preserved in a labeled plastic envelope. A total of 154 samples were collected and were used in this study.

2. Chemicals

Glacial acetic acid, 95% ethanol, 70% ethanol, acrylamide, bisacrylamide, ammonium persulphate, 37% formaldehyde, sodium thiosulfate, sodium dedocyl sulphate (SDS), sodium carbonate, Tris-HCl, EDTA, NaCl, sodium acetate, ethidium bromide, phenol, chloroform, isoamyl alcohol, agarose, urea, NaOH, detergent, silver nitrate, tris base, boric acid, and TEMED (N, N N' N' tetramethylethylene diamine).

3. Reagents

Digestion buffer (1M Tris HCl pH 8.0, 0.5M EDTA, 10% SDS, 5M NaCl), 3M sodium acetate, 2M sodium acetate, TE buffer (pH 8.0), 10% NaOH, 0.5% acetic acid in ethanol, chloroform – isoamyl alcohol (24:1), 20X SSC, Proteinase K (20 μg/μL), silver nitrate staining solution, fix/stop solution, 0.5X TBE buffer, 6% acrylamide, 10X TBE buffer, 10% ammonium persulphate, bind silane, rain-X, 40% acrylamide, Multiplex kits: CTT Multiplex (CSF1PO, TPOX, TH01), Taq DNA

polymerase (5 μ/μL), STR 10X buffer (500mM KCl, 100mM Tris-HCl pH 9.0), STR 10X primer, K562 DNA, DNA markers, STR ladder and STR 2X loading solution.

4. Apparatus

Water bath, 1.5 mL micro centrifuge tubes, micro centrifuge, vortex mixer, volume pipette (1000 μL, 200 μL, 100 μL, 20 μL and 10 μL), parafilm, monopan balance sensitive for smaller weight, forceps, thermal cycler, ice, microwave, freezer, refrigerator, fume hood, gel tray, container, gel comb, electrophoresis gel box, agarose gel electrophoresis, UV transilluminator (302 nm), polyacrylamide gel electrophoresis, glass plates, side spacers, 14 cm vinyl sharkstooth comb, power supply, clamps, pen marker, sponges, tissue paper, filter paper, plastic wrap, syringe and needle, plastic tray, 0.2 mL or 0.5 mL microcentrifuge tubes (compatible with thermal cycler)

REAGENT PREPARATIONS AND METHODOLOGY

1. DNA extraction from buccal swab (saliva)

Reagent preparations

1M Tris HCl pH 8.0

A total of 121.1 g of Tris-base were added into 800 mL of deionized water. The pH was adjusted to pH 8.0 by adding HCl and then made it into 1000 mL and autoclaved.

0.5M EDTA pH 8.0

A total of 186.1 g of Na_2EDTA was added in 800 mL of deionized water with vigorous stirring. The pH was adjusted to pH 8.0 with NaOH pellet. Then, the final volume was adjusted to 1000 mL by adding of deionized water and autoclaved.

10% SDS

A total of 10 g of SDS were added in 90 mL of deionized water in an autoclaved sterile bottle. It was heated to assist dissolution and then adjusted to pH 7.2 by adding a few drops of HCl and the final volume was adjusted to 100 mL.

5M NaCl

A total of 292.2 g of NaCl was added to 800 mL deionized water and made up to 1000 mL.

Digestion buffer

A total of 1.0 mL of 1M Tris HCl pH 8.0 was added together with 2.0 mL of 0.5M EDTA, 20.0 mL of 10% SDS and 1.0 mL 5M NaCl were added along with deionized water to make it to 100 mL. The buffer was then autoclaved and stored at room temperature.

Proteinase K (10 mg/mL)

A total of 50 mg of proteinase K was added to 5 mL of sterile deionized water.

Chloroform-isoamyl alcohol (24:1)

The buffered phenol was combined with an equal volume of a 24:1 chloroform: isoamyl alcohol solution. Then the solution was covered with 0.01M Tris HCl (pH 7.5) and store at 4°C in a brown glass bottle.

70% ethanol

An amount of 350 mL of absolute ethanol was mixed with 150 mL of deionized water.

2M sodium acetate

An amount of 200 mL of 3M sodium acetate was added with 100 mL deionized water.

3M sodium acetate

A total of 102.025 g of sodium acetate was dissolved in 200 mL deionized water; pH was adjusted to 5.2 with glacial acetic acid. The final volume was adjusted to 250 mL and autoclaved.

TE buffer

An amount of 10 mL of 1M Tris HCl was mixed together with 0.2 mL of 0.5M EDTA and added with 989.8 mL of deionized water. The buffer was aliquot in 100 mL buffer and autoclaved.

DNA extraction methodology

In order to prevent contamination all procedures was carried out wearing gloves. Both sides of cotton swab was cut and placed in a 1.5 mL microcentrifuge tube with the same label as the label of samples to avoid mistaken identity of the samples. A total of 500 μL of digestion buffer was added in each sample to lyse cell walls and plasma membrane. An amount of 12 μL of Proteinase K (20 mg/mL) was added to break protein component of cells in the sample. The sample was incubated overnight at 37°C.

Next day 120 μL buffered phenol was added to each sample to precipitate protein. The contents of tube were mixed vigorously and centrifuged at 10000 rpm for 3 minutes. The supernatant was transferred to a new 1.5 mL microcentrifuge