UNIVERSITI SAINS MALAYSIA

UNIVERSITI SAINS MALAYSIA

DISTRIBUTION OF ALLELE FREQUENCIES OF SHORT TANDEM REPEATS (STRs), D16S539, D13S317 AND D7S820 IN KADAZAN POPULATION GROUP IN SABAH, MALAYSIA

by

GABRIEL LOH ONN KIT

Dissertation submitted in partial fulfillment
of the requirements for the degree
of Bachelor of Health Sciences (Forensic Science)

2008

CONTENTS:

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENT:

The present study is "Polymorphism of STRs D7S820, D13S317 and D16S539". This thesis could not have been prepared without the assistance of a large number of people. First and foremost, I would like to take this opportunity to thank my research supervisor, Mr.Panneerchelvam who helped me a lot in completing this thesis paper because he gave a lot of references and guidance so as enabling me to complete the present study. In addition, his careful explanation strengthened the paper I also wish to show my deep appreciation to my co-supervisor, Professor Dr. Norazmi Mohd. Noor for spending his time to figure out this research.

Doing this study would have been impossible without the expert guidance provided by many people especially Mrs. Linda for the technical assistant, guidance and suggestions in the laboratory. I should thank to Mrs Rosniah and Mrs Hafizah for providing me the material to complete my thesis too. I express my deep appreciation to all of them. Besides that, I would like to show my appreciation to my friends, Ms Wahidah and Ms. Allia who had given me a helping hand in doing the laboratory work. I am thankful to my friends, Mr. Reynold and Ms. Joanne who helped me in collecting buccal swab samples from Kadazan population in Sabah

Finally, I would like to thank once again for those who had helped me in doing this thesis paper. Their co-operation and generous courtesies at all times made the completion of this study a pleasant experience. Writing this paper gives me a lot of experience and practice that will be useful in my future life.

ABSTRACT:

Polymerase Chain Reaction (PCR) based Short Tandem Repeat (STR) DNA profiling has revolutionized forensic human identification and individualization. The STR DNA profiling provides certainty in human identification because of the markers having high heterozygosities and discrimination power probability. The distribution of STR alleles utilize in DNA profiling that is between populations. For application of STR DNA profiling in kinship testing, it is essential to have a population database for individual ethnic groups. With this objective in this research study, the population database for three validated STR namely D7S820, D13S317 and D16S539 for Kadazan population Eastern Malaysia was compiled. The heterozygosities (H) for these three STR are D7S820 (0.9136), D13S317 (0.9152) and D16S539 (0.9283). Power of discrimination (PD) for these three STR are D7S820 (0.9226), D13S317 (0.9152) and D16S539 (0.9283). The cumulative discrimination power (CDP) for these three STR for Kadazan population is 0.9995. This database can be used in criminal cases and kinship testing in Kadazan population involving individual.

INTRODUCTION:

Forensic science is the application of science to the requirement of law. There are two principles used in forensic science, Locard's Exchange principle and Individualisation principle. The ability to individualize a person is very important to bring the culprits to the court. The ability to type DNA from biological evidence is an important development in forensic science since the advent of fingerprint analysis (Angel and Paula, 2005). DNA technology enables the forensic scientist to exonerate innocent individuals and to reduce the number of potential contributors to few individuals (Angel and Paula, 2005).

Since the discovery of DNA profiling in 1985, forensic genetics has experienced a technical revolution, both in the type of DNA markers used and in the methodologies or its detection. DNA profiling was first described in 1985 by Sir Alec Jeffrey and his colleagues (Jeffrey *et.al*, 1985). It has great value in forensic genetics. Prior to DNA based technology, all forensic genetic casework like paternity testing and criminal casework were performed using classical serological genetic markers, blood groups, human leukocyte antigen, polymorphic protein and enzymes were used. In fact, these tests have limitation when analysis should be done on minimal or degraded material. Besides that, it was a need to get as much information as possible because it is difficult to analyze biological material other than blood and the polymorphic proteins and enzymes as they were infrequent and in low quantity.

Sophisticated electrophoretic methods were introduced to fulfil the forensic genetics requirements. Despite these methods, the information that the forensic geneticists were able to find in many cases was insufficient (Angel and Paula, 2005). DNA typing has advantages over traditional method, the first of which being it is more conclusive and informative. Analysis can be done in minute or degraded biological sample because DNA is more resistant to degradation than proteins. DNA in an individual is similar and found

in any tissue, so that DNA genotype can be obtained from any tissue whereas the protein markers are restricted to cells where these proteins are expressed (Angel and Paula, 2005).

Genetic typing of polymorphic microsatellite loci (short tandem repeats or STRs) become an effective tool in forensic stain typing from crime cases as well as in identification and paternity cases (Edwards *et.al*, 1991, 1992). Their importance arises from the fact that they are the most informative genetic markers giving high statistical discrimination and individualization (Edwards *et.al*, 1991; Edwards *et.al*, 1992; Hochmeister *et.al*, 1991; Lins *et.al*, 1998). The importance of STR is widely acknowledged and documented due to its extensive use in forensic, medical and ethnogenetic fields for proper utilization of their discrimination power and polymorphic nature in human identification, medical diagnosis and linkage studies (Lin *et.al*, 1998; Ban *et.al*, 2001; Alonso *et.al*, 2003; Smyth *et.al*, 1996 and Reato *et.al*, 1998). STR loci are associated with high heterozygosities in the human population with differences in allele frequencies and heterozygosities between population groups (Wall *et.al*, 1993). STRs are expected to escape the allele frequency distortions observed at loci that were ascertained on the basis of polymorphism in human populations (Rogers *et.al*, 1996).

Short tandem repeats loci consist of di- to penta-meric repeats with fragment length smaller than 300 bp (Jeffreys *et.al*, 1985 and Edwards *et.al*, 1991). They are named according to the length of core repeat units. Dinucleotide refers to those DNA sequence with two core repeat units while trinucleotide for three core repeat units and so on. Tetranucleotide repeat loci are preferred due to the lower amount of "stutter" produced during PCR. Stutter products are additional peaks that can complicate the interpretation of DNA mixtures by appearing in front of regular allele peaks. The number of short tandem repeats loci can reach up to 10^5 per genome (Krawezak and Schmidtke, 1998). The analysis of STR polymorphism by PCR-based method offers few advantages over RFLP

typing (Restriction Fragment Length Polymorphism-typing): (1) STR loci can be typed with a high degree of specificity and sensitivity in a short time period, (2) these loci can be amplified easily even in degraded biological materials and (3) typing of multiple loci can be accomplished in a single multiplex reaction (Hochmeister *et.al*, 1991 and Lins *et.al*, 1996).

The development of Polymerase Chain Reaction (PCR) technique (Mullis, 1985) has enabled the powerful means for isolation and analysis of small portions of very large genomes. STR loci can be amplified by this technique (Weber & May, 1989; Edwards *et.al* 1992). The throughput of STR analysis can be increased by multiplexing STR loci together. The most effective way is to amplify several loci simultaneously in one reaction followed by electrophoresis in a single lane of a gel (David *et.al*, 1997).

Short tandem repeats can be divided into three categories which are simple repeats, compound repeats and complex repeats. Simple repeats are those core repeats with same length and sequence. Compound repeats are made up of two or more simple repeats while complex repeats contain several core repeats blocks of variable unit length as well as variable intervening sequences (Urquhart *et.al*, 1994). The present study focuses mainly in three STR loci which are known as D16S539, D7S820 and D13S317. They are all tetranucleotide repeats. The D16S539 locus is located in chromosome at 16q24 – qter. The locus consists of a variable number of tetrameric AGAT motifs, repeated from 8 to 15 times and giving a total of 8 common alleles. On other hand, D7S820 locus is situated at 7q11.21 to 22. The core repeat unit of this locus is AGAT and the numbers of known alleles are 9. Repeat sequences represent all four possible permutations (like AGAT is used for AGAT, GATA, ATAG or TAGA). The first alphabetic representation of the repeat (AGAT) is employed according to the precedent of Edward *et.al* (1992). The D13S317 locus is located at 13q22 – q31. Its repeat sequence is AGAT. The numbers of

known alleles are 9 (Butler, 2001; GenePrint STR System Technical Manual D004, Promega Corporations, Madison).

Compilation of empirical data in a population is essential and a prerequisite for application in forensic case works. Hence in this study, a database on the three validated STRs which are D16S539, D7S820 and D13S317 were compiled for 102 unrelated healthy individuals of Kadazan population from Sabah.

LITERATURE REVIEW:

Chromosomes carry information relating human inheritance. Every single human cell consists of 23 pairs (2n) of chromosomes (Tjio and Levan, 1956) except for red blood cells. Chromosome can be visualized during metaphase stage of mitosis. Each contains a condensed or constricted region called the centromere which establishes the general appearance of each chromosome. Extending from either side of the centromere are the arms of the chromosome. Depending on the location of the centromere, different arm ratios are produced (Snustad & Simmon, 2006). Chromosomes are divided into metacentric, submetacentric, acrocentric or telocentric based on the position of centromere. Short arm of chromosome is designated as "p" arm (p stands for "petite") while the longer arm is known as "q" arm.

The haploid number (n) of chromosomes is equal to one-half the diploid number. The total set of genes contained in a haploid set of chromosomes constitute the genome of the species. Homologous pairs consist of identical gene sites along their lengths which is called as locus. They are identical in their genetic potential. In sexually reproducing organisms, one member of each pair is obtained maternally and one is obtained paternally (Snustad & Simmon, 2006). Therefore, each diploid organism inherits one gene from mother and one gene from father as a consequence of biparental inheritance (Snustad & Simmon, 2006). Sex-determining chromosomes are often not homologous in size, centromere placement, arm ratio or genetic content. Males have one Y chromosome and one X chromosome while females carry two homologous X chromosomes. X and Y chromosomes are not strictly homologous. Y is the smallest chromosome in human karotype (Snustad & Simmon, 2006)

DNA or deoxyribonucleic acid is a nucleic acid and its buildings blocks are called as nucleotides. DNA consists of a nitrogenous base, a pentose sugar (5-carbon sugar) and

a phosphate group (Snustad & Simmon, 2006). Nitrogenous base is divided into two kinds which are nine-member double-ring purines and six-member single ring pyrimidines. Two types of purines (adenine and guanine abbreviated A and G) and three types of pyrimidines (cytosine, thymine and uracil abbreviated C, T and U) are found in nucleic acids. Nucleoside is a term which describes a molecule composing a purine or pyrimidine base and a ribose or deoxyribose sugar. If a phosphate group is added to the nucleoside, the molecule is known as nucleotide (Snustad & Simmon, 2006).

Watson and Crick had proposed the structure of DNA in 1953. They have suggested few points based on DNA structure: (1) two long polynucleotide chains around a central axis forming a right handed double helix, (2) the two chains are antiparallel, (3) the bases of both chains are flat structures lying perpendicular to the axis and they are stacked on one another, (4) nitrogenous bases of opposite chains are paired to one another by the formation of hydrogen bonds, only A = T and G ≡ C pairs are allowed, (5) each complete turn of the helix is 34 Å (3.4 nm) long; hence, 10 bases exist per turn in each chain, (6) in any segment of the DNA molecule, alternating major grooves and minor grooves are along the axis and (7) the double helix measures 2.0 nm in diameter (Snustad & Simmon, 2006). Base pairing has revealed that amounts of A equalled T and that G equalled C (Snustad & Simmon, 2006). The specific A = T and G ≡ C base pairing is the basis for the concept of complementarity which describes the chemical affinity provided by the hydrogen bonds between the bases.

DNA is considered a polymer with millions of nucleotides. The largest human chromosome, chromosome number 1 is 220 million base pairs long (Gregory *et.al* 2006). Sense strand is a DNA sequence if its sequence is the same as that of a messenger RNA copy that is translated into protein. Antisense strand is complementary to the sense sequence. In other word, antisense strand becomes the template for producing the

messenger RNA. Both sense and antisense sequences can exist on different parts of the same strand of DNA (Munroe, 2004). DNA is present in highly twisted condition called DNA supercoiling. When DNA is in relaxed state, a strand usually circles the axis of the double helix once every 10.4 base pairs (Benham and Mielke, 2005). Positive supercoiling is a term to describe the DNA is twisted in the direction of the helix and the bases are held more tightly together while negative supercoiling is when the DNA is twisted in the opposite direction and the bases come apart more easily. Naturally, most DNA has slight negative supercoiling that is introduced by topoisomerases (Champoux, 2001). These enzymes are needed during transcription and DNA replication process (Wang, 2002).

When denaturation of double stranded DNA occurs, the hydrogen bonds will break down while no covalent bonds break. Strand separation can be induced by heat or chemical treatment. When DNA strand is separated, the viscosity of DNA decreases and both the ultraviolet (uv) absorption and the buoyant density increases. This increase in uv absorption is known as hyperchromic shift (Snustad & Simmon, 2006). The property of denaturation-renaturation of nucleic acids forms the basis for one of the most powerful techniques in molecular genetics which is called molecular hybridization. Molecular hybridization techniques have brought a massive change in genetic world. Hybridization can occur in solution or when DNA is bound either to a gel or to a specialized binding filter which is known as DNA blotting procedure whereby hybridization serves as a way to "probe" for complementary nucleic acid sequences (Snustad & Simmon, 2006).

In the approximately 3 billion base pairs of DNA of the human genome are an estimated 35,000 genes (The International Human Genome Mapping Consortium, 2001). All human genes are encoded in roughly 10% of the human genome. The remaining 90% of human genome represents "noncoding" parts of the genome because they do not contain any genetic information relevant for protein synthesis. Genetic variation in coding DNA is

limited with the exception of the Human Leukocyte Antigen (HLA). This is due to that expressed genes are subjected to selection pressure during evolution to maintain their function. In contrary, the noncoding region is subjected to selection pressure and thus mutations in these regions are always kept and passed down to next generation, giving to a tremendous increase in genetic variability. This region is very informative and useful for identification purposes (Angel and Paula, 2005). About 30% of the noncoding DNA consists of repetitive sequences which are divided into tandemly repetitive sequences and interspersed elements two classes. Majority of forensic typing systems nowadays are mainly based on genetic loci with tandem repetitive DNA sequences (Jeffreys *et.al*, 1985).

Tandemly repeated regions can be divided into minisatellites (Jeffreys *et.al*, 1985) and microsatellites or short tandem repeat (Litt and Luty, 1989; Tautz, 1989). Minisatellites is also known as variable number tandem repeats (Nakamura *et.al*, 1987). Minisatellites consist of repeat units vary from 9 to 100 bp in length reiterated tandemly for a total length of 500 bp to 20 kb while microsatellite (STR) is composed of sequence motifs ranging from 2 to 7 bp for a total length between 50 and 500 bp (Angel and Paula, 2005). STRs are distributed throughout the human genome occurring with a frequency of one locus every 6-10 kb (Beckman and Weber, 1992). Unequal crossing over and gene conversion (Jeffreys *et.al*, 1985) are involved in the variability of minisatellites while variability in microsatellites is caused by replication slippage (Beckman and Weber, 1992). Genetic variation between individuals is mainly based on these minisatellites and microsatellites systems but it is also based on differences in the DNA sequence itself because the repeats can have slight differences in the sequence (Angel and Paula, 2005).

STRs can range from the extremely complex STRs to simple STRs (Urquhart, 1994). Complex STRs have more variability while simple STRs are easy for standardization and have low mutation rates. There are few parameters to be considered

when selecting ideal STRs which include artifactual bands, the robustness and the size. Generally, short sizes are preferred because the size of the amplified product is critical in degraded samples and small fragments can easily be amplified compared to large fragments when degradation of DNA happened (Urquhart, 1994).

Stutter peaks are the most common artefacts in PCR STR analysis. It is caused by slippage of the Taq polymerase enzyme during copying of the STR allele. This problem happens when STR loci are co-amplified in a multiplexed system and is a normal consequence of amplification process which is not optimal for all of the constituent loci. The frequency of stuttering depends on loci, multiplex system in use and alleles within the locus. Stutter bands have smaller peak area compared to main band; normally 15% or less of the peak area of the main band (Gill *et.al*, 1985).

Non-specific also artefacts occured in STR profiling and which are caused by non-specific priming in a multiplex system (Gill *et.al*, 1997). It happens frequently when more loci are co-amplified or when the DNA is degraded (Gill *et. al*, 1997). Most of the artefacts encountered have low peak areas, aberrant peak morphology, do not fall within the allelic range of the locus or loci with the appropriate coloured fluorescent dye and the position of artefact bands are seldom consistent with band shift expectations (Gill *et.al*, 1997).

Chromosomal abnormalities also contribute to STR extra peaks in STR data but it is rare to happen. Chromosomal abnormalities include chromosomal translocation, somatic mutations and trisomy. It causes only minor problem because the same pattern of DNA bands will be present in both the crime sample and the reference sample from the matching suspect (Gill *et.al*, 1997). "N" bands causes the addition of a single nucleotide to the terminus of a newly synthesized DNA molecule. This can result the formation of two species of DNA molecule generated from the same target sequence and only differ in size

by one base (Gill *et.al*, 1997). This is due to not all copies (N fragments) will be added another base to form 'N+1' fragments. Incomplete conversion of 'N' to 'N+1' can happen too at some loci but not others (Gill *et.al*, 1997).

Due to the advance of the technology, the speed of analysis has been improved for forensic DNA analysis. Forensic scientists can obtain the result within few hours. PCR is a technique for the in-vitro amplification of specific DNA sequences by the simultaneous primer extension of complementary strands of DNA (Mullis *et.al*, 1985). PCR uses two primers (forward and reverse primers), each complementary to opposite strands of the region of DNA that have been denatured by heating. The primers are arranged so that each primer extension reaction directs the synthesis of DNA toward the other. PCR consists of three steps which are denaturation, annealing and extension steps. PCR requires several components (Mullis, 1985) which include template DNA, two primers, Taq polymerase to synthesis the region to be amplified, dNTPs, buffer solution, divalent cation and monovalent cation like potassium ions.

Automated temperature cycler or also called as thermocycler allows control of these three steps successively. All these three steps constitute as one cycle and PCR is carried up to 30 cycles. As a result, million copies of DNA of interest can be produced (Mullis, 1985). The minisatellite D1S80 (pMCT118) was the first to be used in PCR amplification based DNA typing (Beckman and Weber, 1992).

Manual electrophoretic systems were firstly used to analyze STRs. Denaturing polyacrylamide gels are recommended for standardization purposes. Use of fluorescent-based technology and DNA sequencer has revolutionized forensic DNA profiling by allowing the typing of large multiplexes and the automation of the procedure. The main advantage of the use of sequencers is automation and the possibility of using intelligent systems of interpretation. Sequence reference allelic ladder is important in STR typing. It

comprises most of the alleles of the STR system occurring in population (Gill *et.al*, 1997; Gill *et.al*, 2000).

Successful DNA profiling of forensic samples lies entirely on the quality and quantity of DNA that is recovered from the specimen. This is of importance when samples are to be analyzed having suffered from environmental stress. Trace compounds may be co-extracted which can influence the quality of the extracted DNA. Thus, the efficiency and sensitivity of the extraction procedures are important considerations in selecting appropriate extraction method. Phenol-chloroform method (Sambrook *et.al*, 1989) is a well-established extraction method. Though this method is time-consuming and involves toxic reagents, it is applicable for samples containing only very little amounts of DNA like hair shafts or samples suspected containing PCR-inhibiting substances (Silvano *et.al*, 2005). This method is applicable to wide range of cell types and stain materials (Silvano *et.al*, 2005).

Differential extraction should be performed in cases of mixed samples of the stain material consists of mixture of sperm and nonsperm cells (Gill *et.al*, 1985). Differential extraction is used to isolate the sperms from other cells before conducting organic extraction. For the extraction of DNA from hairs, buffer system containing Proteinase K and Ca^{2+} is used instead of EDTA to enhance the efficiency of DNA extraction (Hellman *et.al*, 2001).

DNA quantification method is performed to ensure the efficiency of the extraction step. This step is done to estimate the amount of DNA extracted before added to the PCR master mix to avoid excess of DNA template. Photometric/fluorometric determination of the DNA amount is common quantification techniques. A negative quantitation result does not show absence of DNA. The actual DNA amount may be masked as a result of substances included in the extract that interfere with the detection method. As a

consequence, some laboratories refrain from quantitation per se and directly apply an aliquot of the extract to PCR to quantify by analyzing peak heights of the resulting DNA profile (Silvano *et.al*, 2005).

Silver staining is one of the techniques used to detect proteins and DNA separated by gel electrophoresis. It is a very sensitive tool for protein and DNA visualization with a detection level ranging from 0.3 – 10 ng level (Sorensen *et.al*, 2002). The basic mechanism underlying silver staining is that binding of silver ions to the amino acid side chains primarily the sulfhydril and carboxyl groups of proteins (Oakley *et.al*, 1980; Merril *et.al*, 1981; Merril *et.al*, 1986). Subsequently, the metallic silver is reduced (Oakley *et.al*, 1980; Merril *et.al*, 1981; Merril *et.al*, 1986). DNA bands or protein bands will be visualised as reduction occurred. A number of alteration in the silver staining procedure can shift the oxidation-reduction equilibrium in a way separated proteins or DNA will be visualized either as positively or negatively stained bands (Merril *et.al*, 1986).

Based on its protocol, silver staining can be divided into two groups: 1) silver amine or alkaline method, and 2) silver nitrate or acidic methods (Merril *et.al*, 1986). Silver amine method has lower background and is more sensitive method than silver nitrate method but it requires longer time to produce the result. Silver nitrate method can give result in short of time but it is not as sensitive as silver amine method (Sorensen *et.al*, 2002). Protocols of silver staining has been modified with the introducing of either glutaraldehyde or formaldehyde based sensitizes in the fixing and sensitization step allowing the chemical modifications into proteins but the use of these chemicals have caused the cross-linking of two lysine residues within protein chain (Sorensen *et. al*, 2002).

Kadazan is an native population living mainly in the state of Sabah in Malaysia especially on the west coast of Sabah. The term "Kadazan" was believed a political derivative that came into existence in the late 1950s to early 1960s. There is no proper

historical record exists pertaining to the origins of the term or its originator (Tunggolou, 2004). The term originates from the word "kakadazan" (towns) or "kedai" (shop) and from the claim of Kadazan politicians, Datuk Peter J. Mojuntin. Mojority of Kadazans are Christians mainly Roman Catholics and some Protestants. Islam is also one of the religions for Kadazan but in much smaller number (Tunggolou, 2004). Sumazau is the popular dance among Kadazans. It is performed by male and femal adults during joyous ceremonies and special occasion like wedding feasts (Mercurio, 2006; Matusky, 1985).

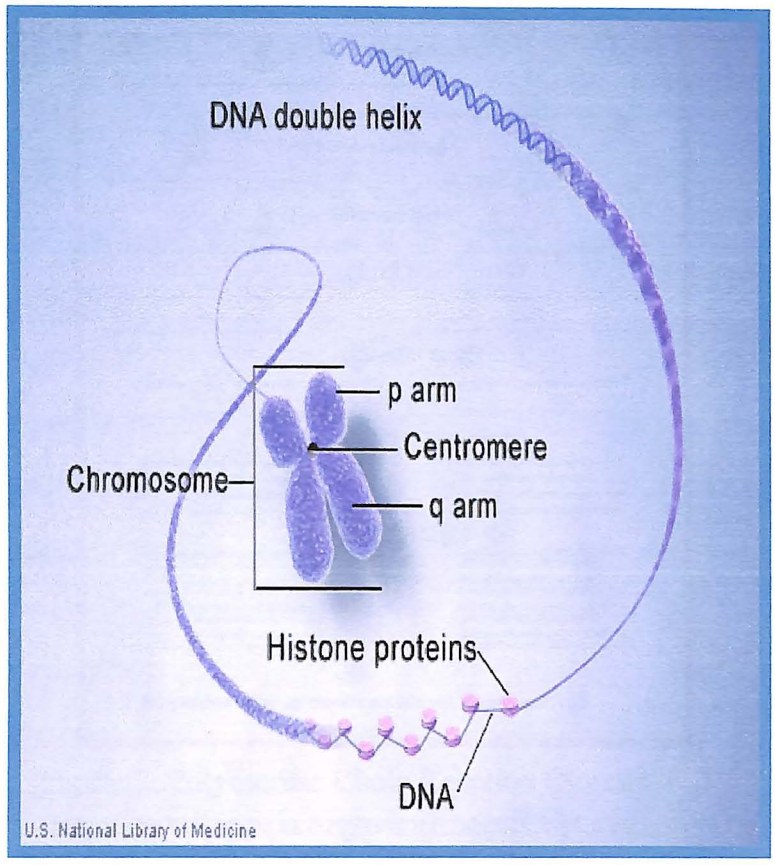

Figure 1: Human Chromosome (Source:

http://ghr.nlm.nih.gov/handbook/illustrations/chromosomestructure.jpg

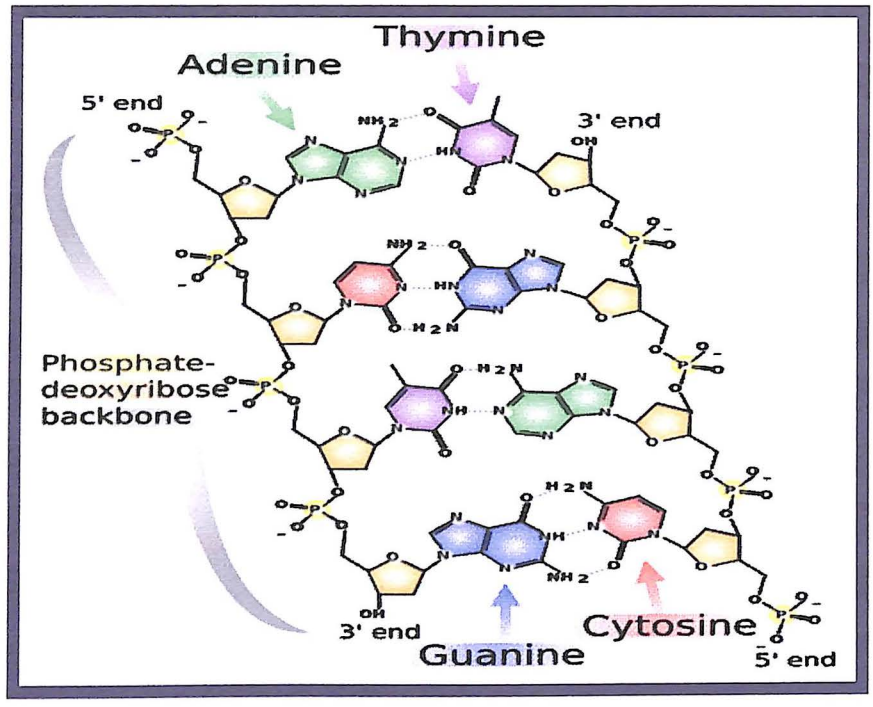

Figure 2: Chemical Structure of DNA (Source:

http://en.wikipedia.org/wiki/Image:DNA_chemical_structure.svg)

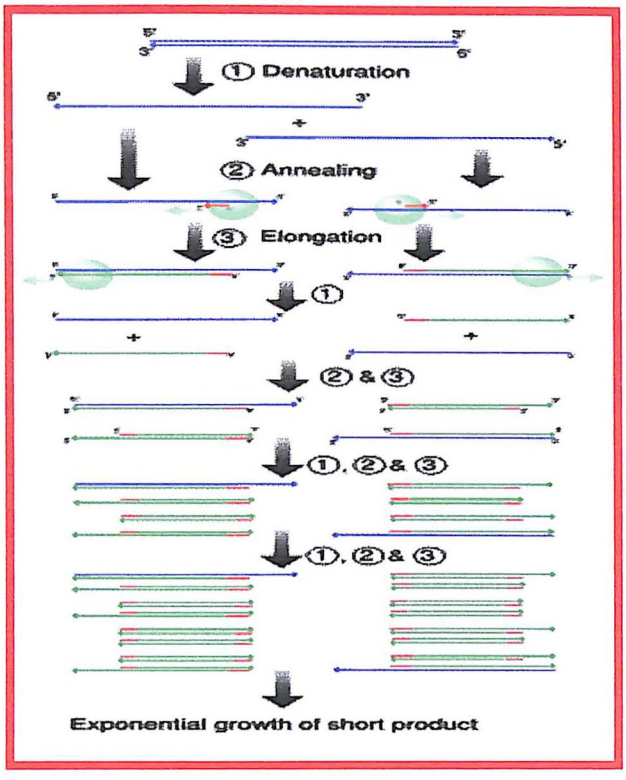

Figure 3: Polymerase Chain Reaction (Source: http://en.wikipedia.org/wiki/Image:PCR.svg)

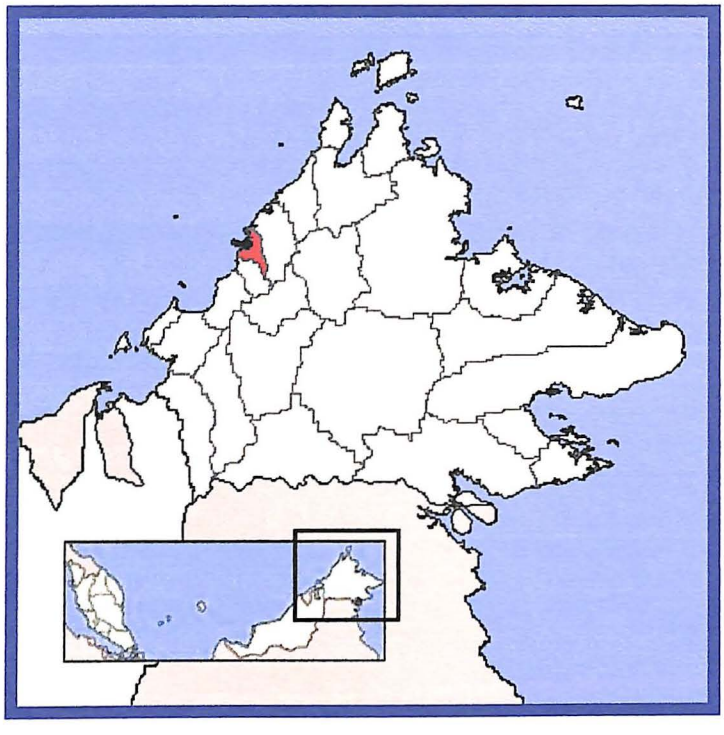

Figure 4:Location of Sabah

(Source:http://en.wikipedia.org/wiki/Image:DynamicSabahblankcolorschemeKK.PNG)

OBJECTIVES OF THE STUDY

GENERAL OBJECTIVES:

1. To understand the principle of manual DNA typing.
2. To familiar DNA typing methods.

SPECIFIC OBJECTIVES:

1. To compile a population data base on three validated STRs D16S539, D7S820 and D13S317 for Kadazan population.
2. To calculate the allele frequency for the three validated STRs in Kadazan population.

MATERIALS AND METHODS:

1.0 Materials

All equipments, reagents and materials used in this study were sterilized prior to use to avoid contamination. The STR typing was performed according to the guidelines provided by the manufacturer of STR kits (GenePrint STR System Technical Manual D004, Promega, USA).

1.1 SAMPLE SOURCE:

Buccal swabs were collected from 102 healthy unrelated Kadazan individuals of Malaysia. Prior to sample collection, an informed consent was obtained from each subject. Sterile cotton buds with two sides were used and each participant was asked to streak one side of the cotton bud from left cheek for 10 seconds and another side of cotton bud from right cheek for another 10 seconds. The cotton buds were dried at room temperature without exposing to direct sunlight to avoid the degradation of the samples. The samples were kept in zipper envelope (9cm X 15cm) and labelled with appropriate information like subject's name, age, sex and address.

1.2 REAGENTS:

Digestion buffer (1M Tris HCL pH 7.5, 0.5M EDTA, 20% SDS, 5M NaCl), proteinase K (20μg/μl) (Promega, USA), Chloroform-isoamyl alcohol (24:1), 3M sodium acetate, 2M sodium acetate, 70% ethanol, TE buffer, 10% NaOH, 0.5% acetic acid in ethanol, bind silane (Promega, USA), 10% ammonium persulphate, 40% acrylamide, 10X TBE buffer, 0.5X TBE buffer, CTT Multiplex kit consisting of STR 10X buffer (500 mM KCL, 100 mM Tris-HCl with pH 9, 15 mM $MgCl_2$, 1% Triton X-100, 2 mM each dNTP), multiplex 10X CTT primer pair mix, Taq DNA polymerase (5μg/μl), STR 2X

Loading Solution (10 mM NaOH, 95% formamide, 0.05% bromophenol blue, 0.05% xylene cyanol FF), CTT Allelic Ladder Mix (Promega, USA), acylamide solution, silver nitrate staining solution, fix/stop solution, gel developer, Rain X (Blue Coral-Slick 50, USA).

1.3 CHEMICALS:

Tris base (Promege, USA), concentrated HCl, Na_2EDTA (Promega, USA), NaOH (Merck, Germany), NaCl (Merck, Germany), sodium dodecyl sulphate (SDS) (Bio-Rad Lab), chloroform (Merck, Germany), isoamyl alcohol (Merck, Germany), sodium acetate (Merck, Germany), glacial acetic acid (Merck, Germany), absolute ethanol (Merck, Germany), EDTA (Promega, USA), ammonium persulphate (Promega, USA), acrylamide (Promega, USA), bisacrylamide (Promega, USA), boric acid (Promega, USA), urea (Promega, USA), silver nitrate (Promega, nitrate), 37% formaldehyde (Promega, USA), sodium thiosulphate (Promega, USA), sodium carbonate (Promega, USA) and TEMED (N,N,N',N' tetramethylethylene diamine) (Promega, USA), phenol (PIERCE, USA), agarose powder (Promega, USA).

1.4 APPARATUS:

Desiccators Nucerite (Nalgene/Syborn Corp), Vortex mix EVM-6000 (ERLA), microfuge 16M (National Labnet Co.), Gilson varying pipette (1000µl, 200µl, 100µl, 20µl and 10µl), parafilm (American National Can), Biometra Gene Ray-UV photometer, Agarose gel apparatus model MGU-202T (C.B.S Scientific Co., California), Electrophoresis power supply (Amersham Pharmacia, Biotech, USA), MJ Research PTC-200 Peltier Thermal Cycler (Gradient cycler) used for 0.2µl tubes, MJ Research PTC-100 Peltier Thermal Cycler used for 0.5µl tubes, monopan balance

sensitive for smaller weights (DRAGON 204 Mettler Toledo), Fischer & Pagkal N308 Freezer, Hitachi refrigerator, Mammart waterbath, SA 32 Electrophoresis apparatus (GIBCO BRL Sequencing System), High voltage powerpack (EC 3000-90), fumehood (Model: RICO), plastic trays for staining, stopwatch.

1.5 BUCCAL SWAB COLLECTION:

Buccal swabs were taken from 102 healthy unrelated random Kadazan 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, labelled with details of collection. The samples were kept in dry environment at room temperature.

2.0 METHODS

2.1 Reagent preparations

1M Tris HCl pH 7.5

121.1g Tris base is diluted in 800 ml deionised water and the pH is adjusted to 7.5 with concentrated HCl. Solution is topped up to 1000ml and autoclaved.

0.5M EDTA

186.1g Na_2EDTA is diluted in 800ml deionised water and the pH is monitored to 8.0 with NaOH pellets. The solution is made up to 1000ml and autoclaved.

5M NaCl

292.2g NaCl is added to 800ml deionised water and made up to 1000ml.

20% SDS

An empty bottle is autoclaved. 100g Sodium dodecyl sulphate is added to 400ml deionised water. Solution is stirred with heat using magnetic stirrer. The solution is topped up to 500ml and stored in an autoclave bottle.

Digestion buffer (Do not autoclave)

1ml of 1M Tris HCl pH 7.5, 2ml of 0.5M EDTA, 10ml of 20% SDS, 1ml of 5M NaCl and 86ml of deionised water are mixed together.

Proteinase K

20mg of Proteinase K is dissolved in 1ml of deionised water.

Chloroform-isoamyl alcohol

240ml of chloroform is added to 10ml of isoamyl alcohol

3M sodium acetate

102.025g of sodium acetate is added to 200ml of deionised water. The pH of the solution is adjusted to 5.2 by using glacial acetic acid. Solution is topped up to 250ml and autoclaved.

2M sodium acetate

16ml of 3 sodium acetate is added to 8ml to deionised water.

70% ethanol

350ml of absolute ethanol is mixed with 150ml deionised water.

TE buffer

10ml of 1M Tris HCl is mixed with 0.2ml 0.5M EDTA. 989.8 of deionised water is added and autoclaved.

2.2 DNA EXTRACTION:

Cotton buccal swabs were cut into small pieces and put into a 1.5ml eppendorf tube. A total of 600µl digestion buffer and 12µl Protease K were added into the sample and incubated at 56°C for overnight. 120µl buffered phenol was added and mixed vigorously using the vortex. Samples were centrifuged at 10,000 rpm for 3 minutes. The supernatant was transferred to a new 1.5ml eppendorf tube by using cut tips. A total of 300µl of buffered phenol and 300µl chloroform-isoamyl alcohol were added to the supernatant and the tube was mixed vigorously. The contents were centrifuged again at 10,000 rpm for 3 minutes. By using cut tips, the supernatant was transferred to another new eppendorf tube. One volume of chloroform-isoamyl (same volume as supernatant) was added to the supernatant and vortexed. The contents were centrifuged at 10,000 rpm for 5 minutes. Supernatant from the tube was transferred to a new tube using cut tips. Volume of 500µl chilled ethanol and 50µl 2M sodium acetate were added to the supernatant and the contents inside the tube were mixed by inverting the tube gently. The contents were centrifuged at 10,000 rpm for 10 minutes. The supernatant was discarded while 0.5ml 70% ethanol was added to the tube without the supernatant. The DNA pellet was dislodged and centrifuged at 10,000 rpm for 3 minutes. Supernatant was discarded. The tubes were sealed with parafilm and tiny holes were pinched on the parafilm. The pellets can either be dried by using vacuum for 20 minutes or air-dried in room temperature. A toal of 50µl TE buffer was

added to dissolve DNA pellet and kept overnight at 37°C. The tubes with DNA samples were kept at -20°C for future use.

2.3 QUANTIFICATION OF DNA:

2.3.1 Reagents preparations

Ethidium bromide stock solution

1g of ethidium bromide was dissolved in 100ml of deionized water and stored in an amber coloured bottle with stopper.

1% Agarose gel solution preparation

1g of agarose powder dissolved in 100ml 0.5X TBE. 0.07µl of ethidium bromide stock solution was added to the agarose gel solution.

DNA samples were quantified using spectrophotometer at 260nm. The extracted DNA sample was then diluted to 10 ng/µl using sterile deionized water.

2.4 PCR AMPLIFICATION

Amplification process of three STR loci, D7S820, D13S317 and D16S539 was done according to the recommendations given by the manufacturer (Promega Corporation, Madison, USA). The STR 10X buffer and STRIII 10X Primer pairs were thawed and kept on ice. The number of reactions to be set up was determined. The required quantity of each component of the PCR Master Mix was calculated. The volume per sample is multiplied by the total number of reactions to be carried out in order to get the final volume. The final volume of each reagent was added to a sterile tube, mixed gently and placed on ice. A total of 22.5 µL of PCR master mix was added to each