1.0 INTRODUCTION

A. J. Jeffreys (1985) described the first development of multilocus DNA fingerprints and speculated that these individual-specific DNA patterns are a powerful method for individual identification and paternity testing. In April 1985, the first case, involving a UK immigration dispute, was satisfactorily resolved by DNA fingerprinting (Jeffreys et al., 1985). In addition to parentage testing, DNA fingerprinting has the potential to identify criminals since a person's DNA fingerprint is unique, just like a traditional fingerprint, except for monozygotic twins. Thus, if a criminal leaves some of his cells (in blood, sperm, or hair, for example) at the scene of a crime, the DNA from these cells can identify him. Today, with advance in the DNA fingerprinting technique, DNA typing systems are routinely used in public and commercial forensic laboratories and have replaced conventional protein markers as the methods of choice for solving paternity disputes and criminal cases.

DNA fingerprinting makes use of restriction fragment length polymorphisms (RFLP) due to the presence or absence of restriction enzyme recognition sites, while DNA typing is based on the varying number of tandem repeats in microsatellites, known as short tandem repeats (STR). The main advantage that DNA typing has over DNA fingerprinting is it is sensitive, less technically demanding, requires a very small amount of sample to run compared to the relatively large amount of large molecular weight DNA needed for DNA fingerprinting (Nakamura et al., 1987).

In spite of its accuracy, DNA typing has been effectively challenged in court, most famously in the O. J. Simpson trial in Los Angeles in 1995. The O.J. Simpson trial has brought into sharp focus that DNA fingerprinting and typing have to be

performed very carefully to give meaningful results and the statistics used in analyzing the data have to be beyond question, but in addition, the importance of sample collection and sample custody cannot be over-looked.

The U.S. Crime Act of 1994 and similar legislation in Canada, the United Kingdom and several other countries has paved the way for the development of databases which will hold “DNA fingerprints” for large numbers of individuals previously convicted of violence crimes.

Members of an ethnic group are more related than members of the general population, so data from within that group, rather than from the whole population, should form the baseline data.

The most direct way to gather DNA data usable for identification is for scientists to determine the allele frequencies of commonly used STR loci in ethnic groups in their own countries and the results compiled. It is to this end this project is carried out. The Sikh population in Malaysia has been chosen as my study subject because this is a distinct population in Malaysia but no allele frequency data is available, whereas databases for Malay, Chinese and Indian have already been published (Lim et al, 2001; Seah et al., 2003). For Sikhs in India, there was a paper on allele frequency data (Chattopadhyay et al., 2000).

The objective of my project is to study the pattern and distribution of 9 commonly used STR loci in the Sikhs from Malaysia in order to establish a reference database.

Choosing which STR loci to examine is important for the results to be meaningfully compared to other populations. For this purpose, 9 STR loci (CSF1PO, TPOX, TH01, F13A01, FESFPS, vWA, D16S539, D7S820, and D13S317) which are validated and routinely used in similar studies throughout the world have been selected. Some of these loci are also used in the Combined DNA Indexing System (CODIS) in the United States (Budowle et al., 2000).

I began my project by collecting buccal swabs from unrelated healthy Sikh individuals residing in Malaysia, whereby 119 samples were collected but only 109 DNA samples were acceptable. DNA from each sample was extracted following the standard protocol and amplified using Promega's CTT, FFv, and Silver STRIII Multiplex kits. PCR products were then separated using polyacrylamide gel electrophoresis and silver stained. Results were examined and statistically processed to establish a database of the distribution of these 9 STR loci in Malaysia's Sikh population. This data was then compared to other ethnic groups in Malaysia.

2.0 LITERATURE REVIEW

2.1 Brief History

The foundations of scientific genetics were laid in the second half of the nineteenth century: Charles Darwin recognized the principles of mutation and selection for speciation and evolution; Gregor Mendel discovered that heritable traits split up and recombine in an orderly manner when transmitted to offsprings; Johann Friedrich Miescher first extracted and chemically analyzed nucleic acids; and Wilhelm Roux postulated that chromosomes are the carrier structures of inheritance. The era of molecular genetics, however, did not start until 1944, when Oswald Theodore Avery and co-workers showed that DNA was the genetic material. The following year, 1945, Beadle and Tatum pointed out that one gene encodes one protein, which was then proved incompletely correct, but was very useful in understanding the function of genes. Then in 1953, Watson and Crick demonstrated the famous double helix structure of DNA; 1961, Nirenberg and Matthaei deciphered the genetic code; 1972, Berg and co-workers started molecular cloning of DNA; 1977, Sanger and co-workers developed methods to sequence DNA.

Traditionally, the term ‘fingerprint’ refers to the patterns, which are highly characteristic for any human individual, of the ridged skin of the distal finger phalanges. The term ‘fingerprinting’ has also been used for the electrophoretic and chromatographic characterization of proteins and, more recently, of deoxyribonucleic acid (DNA) molecules. With DNA being the only basis of genetic differences between distinct organisms, DNA fingerprint is the ultimate method of biological individualization.

Genetic uniqueness is brought about by two factors: inheritance and mutation. In diploid organisms, one complement of the genome is inherited from the father and the other from the mother. Random assortment and recombination during meiosis ensure that no offspring has a genome made up of somatic material from only one parent. Genetic diversity is further increased by mutation. Therefore, an ancestral genome will never be reconstituted, even under extreme inbreeding.

Since all genetic differences between individuals are laid down in the primary sequence of their genomic DNA, the most straightforward method of identifying an individual would be to determine this sequence for the genomes under comparison. Using the widely accepted estimate that two homologous chromosomes randomly drawn from the human population differ at a frequency of 1 in 300 bp, sequencing a 15000 bp segment would guarantee that, with 99.9% probability, no pair of unrelated humans living on earth would be found to be identical (Burke et al., 1991).

Individualization by sequencing implies that comparatively long stretches of identical DNA be screened before a difference can be expected to show up. A more efficient strategy is to limit the comparison to regions of the genome which are already known to differ frequently between individuals. Such regions are termed polymorphic sites. Genetic polymorphism is defined as the simultaneous occurrence in the same population of two or more discontinuous variants or genotypes, the frequencies of at least two of the types must be high enough (more than 1% in a given population) that they cannot be accounted for solely by recurrent mutation (Burke et al., 1991).

The first genetic polymorphism, discovered as early as 1900 by Karl Landsteiner, was the ABO blood group system. This polymorphism, along with other

subsequently discovered blood groups and other protein polymorphisms, was analyzed at the level of the gene product and not at that of the gene itself. Analysis at gene level became feasible only after the introduction of molecular techniques in the late 1970s and early 1980s. This new technology opened up a way to study polymorphic traits, namely those defined by the variation between corresponding DNA fragments lengths generated by digestion with restriction enzymes (restriction fragment length polymorphism, RFLP).

The usefulness of genetic polymorphism for the definition of biochemical 'individuality' was realized immediately after the first 'individualities in metabolism' were discovered. The chance of an individual of European descent possessing the most frequent alleles of 15 selected blood groups, serum proteins and red blood cell enzymes is about 1 in 20000 (Payne et al., 2003). With refinements in biochemical, immunological and molecular genetics technology, it is now easily possible to distinguish all living members of the human species at the protein and/or DNA level. Such power of resolution is most effectively achieved by DNA probes (small pieces of DNA that are complementary to the region to be analyzed) that are capable of detecting in any individual a large number of highly polymorphic genetic loci simultaneously. The term 'DNA fingerprinting' was introduced by Alec Jeffreys (1985) to describe the barcode-like DNA fragment pattern generated by such multilocus probes after electrophoretic separation of genomic DNA fragments (Jeffreys et al., 1985).

Subsequently 'DNA profiling' where the combined use of several single-locus probes made the detecting of polymorphism simpler became the preferred method.

2.2 Polymorphisms in the DNA

Methods to individualize have continued to improve since 1900 when Karl Landsteiner first showed that individuals could be placed into different groups on the basis of their blood types. Later, more markers consisting of soluble proteins in plasma were found, making identification test more discriminating. These are serological markers since the detection is based on antigen-antibody reactions (Payne et al., 2003).

Blood proteins are encoded by genes, for example, as there are ten types of phosphoglucomutase (PGM), at least the same number of alternative versions of the PGM gene must be there to produce these ten isoforms. The area of greatest variability in DNA sequences is the non-coding regions of DNA. The human genome consists of approximately 3 billion base pairs of DNA in which there is estimated to be 30,000 genes. The average length of a gene is 5-10 thousand base pairs. This leaves more than 97% of the DNA within the human genome as non-coding. Mutational events and the reduced selection pressure result in greater polymorphisms within these vast regions of non-coding DNA. Approximately 30% of the non-coding DNA is in the form of repetitive sequences and much of this is arranged in tandem repeats. The tandemly repeated sequences have been the focus of forensic use.

DNA markers that distinguished individuals have been known since 1980 with the discovery of a restriction fragment length polymorphism (RFLP) on chromosome 14. When DNA from several individuals was digested with the restriction endonuclease *Eco*R I, separated by agarose gel electrophoresis and detected by Southern blot analysis using a radioactive probe for the D14S1 locus, different banding patterns were seen for different individuals. The alternative forms of the DNA (alleles) which are

characterized by length variations were heritable traits resulting from the presence or absence of a restriction site at this locus.

Repetitive sequence elements which are arranged in tandem are known as satellite, minisatellite, and microsatellite sequences. The three terms refer to different levels of repetition and different repeat lengths (Pena et al., 1993). Satellite DNA is dispersed over almost the entire genome. Satellites, minisatellites and microsatellites can be highly variable and thus form excellent tools for genetic individualization. Their variability is most often due to particular arrays on a given chromosome having different repeat numbers in different people. They form allelic variants and a number of mini- and microsatellites which are highly heterozygous. Polymorphisms created by such elements are termed variable number of tandem repeat (VNTR) polymorphisms.

The nomenclature adopted below follows that of Tautz. In the literature, the use of terminology is not uniform and is sometimes rather confusing. Minisatellites are sometimes equated with VNTRs (Nakamura et al., 1987), but VNTR is a term applicable to all repeat classes. Microsatellites are occasionally referred to as 'simple' sequences or short tandem repeats (STRs), but STR is also used for synthetic tandem repeat probes capable of detecting minisatellites sequences (Tautz and Renz, 1984; Litt and Luty, 1989; Edwards et al., 1991; Vergnaud et al., 1991). Hence, the terms are defined as in the table below:

Type	Degree of repetition (per locus)	Number of loci	Repeat unit length (bp)
Satellite	$10^3 – 10^7$	1 – 2 per chromosome	Two to several thousand
Minisatellite	$2 – 10^3$	Many thousand per genome	9 – 100
Microsatellite	$5 – 10^2$	Up to 10^5 per genome depending on repeat motif	2 – 7

Polymorphism due to variation in the number of elements within a given array is thought to be generated during DNA replication, for example by mutational process of slipped strand mispairing. In addition to allelic variation in repeat number, polymorphism at mini- and microsatellite loci can also be caused by sequence changes in the vicinity of these repeats.

2.3 Variable Number Tandem Repeat Typing

Minisatellites were the first region of DNA to be used in forensic science and paternity studies. These DNA loci range from 500 to more than 20,000 bp. A core unit, typically between 9 and 60 bp, is repeated tandemly along the chromosome. It is the number of repeats that varies between chromosomes and therefore between individuals (Payne et al., 2003). The number of repeats for a particular VNTR locus can be large, with none of the alleles present at a particular high frequency. Alleles at VNTR loci can be separated on agarose gels due to their large sizes and different lengths. After transferring the DNA onto a nylon membrane the size differences can be detected by hybridizing a DNA probe made to the repeat sequence of the alleles. The alleles appear as bands on an autoradiograph. In 1985, Professor Sir Alec Jeffreys is credited with being the first person to use DNA fingerprinting in a criminal case, used a probe that would detect many VNTR loci at the same time (. This process, called multi- locus probing (MLP), was subsequently replaced by single locus probing (SLP) that utilized a probe specific to a single VNTR locus, thereby producing results that were simpler to interpret compared to MLP. VNTR of minisatellites was, however, to be superseded by the advent of the polymerase chain reaction (PCR) used to detect polymorphisms in microsatellites DNA, termed short tandem repeats, STR.

2.4 Short Tandem Repeats

Short tandem repeats (STR) consist of simple tandemly repeated sequence, commonly between 2 and 7 bp in length, which are widely dispersed throughout the genome. STRs are highly abundant: there are as many as half a million STR loci in the human genome, occurring on average every 6 – 10 kb. The STR loci so far analyzed fall into three categories: simple repeats where the repeat elements is repeated in sequence identically; compound repeats comprising two or more simple repeats: and complex repeats comprising several blocks of variable unit length and variable sequence.

The use of STR has many advantages over VNTR. The amount of starting DNA required is considerably less and a high level of degradation can be tolerated. STR typing can be performed on samples with less than 1 ng of degraded DNA, whereas multi- and single-locus probing techniques for VNTR require 50 ng of high molecular weight DNA. The speed of analysis is more rapid in STR typing as there is no requirement for a hybridization process; instead the products of PCR are sized by polyacrylamide electrophoresis.

STR polymorphism observed results almost exclusively from variation in the number of tetra-nucleotide repeats present at the locus, and not from insertion or deletion of one or two bases. Alleles which differ by four bases in these 100-350 base amplification products are easily separated in polyacrylamide denaturing gels. This allows rapid and precise typing. VNTR probes are sometimes not distinctive enough from each other resulting in lack of precision. VNTR bands that appear after probing are sometimes overlapping, causing difficulties in quantifying them.

The discrete nature of alleles of these selected STR loci has also allowed the development of allelic ladders. Allelic ladders are composed of a collection of most or all of the amplified alleles found in the general population. These composites make ideal size markers because the size markers and the amplified unknown alleles will contain not only the same size fragments, but the same sequence fragments. Thus, ladder components and unknowns co-migrate in the gel electrophoresis regardless of the gel matrix or running buffer selected. Consequently, different laboratories using different separation techniques and different detection formats can compare their results with precision and reliability.

Although it is estimated that there are as many as half a million STR loci in the human genome, not all of them are suitable for use in forensic science (Payne et al., 2003). STR loci are chosen when they meet a range of criteria. The primers used in the amplification reaction must anneal to the DNA of all the members of the population, which means that the primers used in amplification will produce PCR products from members of every ethnic group. The selected loci must be highly polymorphic with a high level of heterozygosity. The amplified products must be easily distinguished from one another. This means rejecting markers which contain frequent microvariants (i.e., alleles differing from one another by lengths shorter than the repeat length) as the closer and more random spacing of alleles is more difficult to interpret. Finally, the prevalence of stutter bands (i.e., amplification artifacts which appear one or more repeat lengths above or below the true amplified allele), has led to the rejection of dinucleotide repeats as a class for these applications (Payne et al., 2003; Weber and May, 1989).

An example of a commonly used STR is called vWA, which is found near to the gene sequence for von Willebrand factor. It has the repeat sequence TCTA$(TCTG)_{3-}$

$_4$(TCTA)n. the repeat sequence alters from TCTA to three or four repeats of the sequence TCTG, after which there are a variable number of TCTA repeats. The sequences TCTA and TCTG comprise the repeat unit. The smallest commonly found allele has a total of 11 repeats and the largest allele has a total of 20 repeats. In between there are eight alleles. Either side of the repeat sequences is the flanking DNA. This DNA sequence is largely conserved between individuals and is often the prime site for PCR. The complete length of each PCR product is therefore the length of the flanking DNA on each side of the STR locus, which does not change, plus the variable number of the repeat units. As each individual will have two alleles, being homozygous or heterozygous, the difference in size between the two alleles will be divisible by four.

An example of a complex STR locus is D21S11:

(TCTA)n(TCTG)n[$(TCTA)_3$TA$(TCTA)_3$TCA$(TCTA)_2$TCCATA](TCTA)n

The dinucleotide and trinucleotide insertions greatly increase the discrimination power of this locus, making D21S11 more discriminating than simple STR locus such as vWA. The discrimination power is also increased by the large number of alleles encountered for this locus and this makes it more useful in forensic science.

Highly complex STR loci such as SE33, which comprise tetranucleotide repeats interspersed with a number of dinucleotide sequences are highly discriminating but can be problematical in the definite designation of alleles.

The International Society for Forensic Haemogenetics (ISFH) has provided guidelines for the nomenclature of the loci and alleles. Some of the first STR loci reported were named after the gene sequence to which the locus was nearest. This led to the names such as vWA (von Willebrand factor), THO1 (tyrosine hydrolase), and

F13A1 (coagulation factor XIII A). STR loci without any connection to a protein coding sequence are named after their chromosomal location such as D3S1358 and D19S253, on chromosome 3 and 19, respectively.

Each allele is given a number based upon the number of repeats; therefore THO1 5 has five tetranucleotide repeats. When, as in the case of THO1 9.3, there is an incomplete number of a repeat, the number of individual bases (three in the case of THO1 9.3) is used. Another example of this is the FGA locus, where there are a 22 and a 22.2 allele. The 22.2 allele contains 22 complete tetranucletide repeat sequences and an incomplete repeat sequence containing only two bases. As each person has two alleles, one inherited from the mother and one inherited from the father. A person having a THO1 6 and THO1 9.3 is often typed as THO1 6,9.3. A person who inherits THO1 6 from both parents is THO1 6,6. So far there are 7 alleles reported for THO1 locus, hence yielding 28 genotypes. The relationship between number of alleles, n, and genotypes is given by the formula $(n + 1) n / 2$. As the number of alleles increases, the number of possible genotypes increases to a greater extent.

Using one locus in an STR yields a low level of discrimination. When a second locus is used the level of discrimination is multiplied provided that the two loci are not genetically linked. In forensic cases, between four and 13 STR loci are commonly amplified. Commercially available kits allow up to 10 STR loci to be amplified in one reaction. Thus, genotypes of a sample for all the STR loci examined will be visible and can be recorded as a database.

Forensic application of DNA typing is based on the same principle. Since a considerable proportion of the genetic information embodied in a cell is host-specific

and differs between any two individuals (with the exception of monozygotic twins), analysis of this particular fraction of the genome allows unambiguous identification of any remaining biological trace. As long as DNA can be extracted and analyzed, the chances of success depend on the number and information content of the DNA systems tested. A bigger number of appropriate STR loci tested and more information carried by the loci used in the DNA systems will give more reliable statistical results.

In forensic testing, small samples of somatic tissue (e.g. hair roots, blood, fingernails) are analyzed, whereas in kinship testing the haploid genome conveyed by a single gamete of the alleged represents the issue of concern. Every child inherits one allele from the mother and one from the father. For each STR tested in a child, one allele must have come from the mother and therefore the other STR allele presents in the child must be present in the father's sample. If a child has the genotype THO1 6,7 and the mother is THO1 7,9 then the child must have inherited type 7 from the mother. The biological father must posses the THO1 allele 6. A similar situation must occur for all the STR types tested.

Negative answers to the question of identity are usually quite easy to give; positive identification is much more difficult. Positive identification alone is never a sufficient proof of guilt, DNA typing results add substantially to the body of evidence that a court requires for sound decision making. When two DNA sequences exhibit more differences than are explicable by mutation alone, the suspect is definitely not the source of the material. The only problem here may be to obtain accurate estimates of the mutation rates involved. 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.

3.0 MATERIALS AND METHODS

3.1 Saliva (Buccal) Swab Collection

Buccal swabs were collected from 119 healthy Sikhs (51 females and 68 males) from Kuala Lumpur, Selangor, Kelantan, Melaka and Perak.

Materials

- Cotton buds
- Gloves
- Consent forms
- Small size plastic envelopes (10cm × 4cm)

Methods

Sikh individual was briefed on this project and protocol. Individual who agreed to participate was given a consent form to fill up. Two sticks of buccal swab samples were collected using cotton buds from each individual by swabbing the inner side of the cheek, self-collected or done by the collector with gloved hand. The cotton buds were dried at room temperature then stored in labeled plastic envelope. The stored samples were kept in paper envelope at room temperature.

3.2 Buccal Swab DNA Extraction

Materials

(Solution compositions are in Appendix)

- 1M Tris-HCl pH 7.5
- 0.5M EDTA pH 8.0
- 5M NaCl
- 20% SDS
- Digestion buffer
- Proteinase K (20 mg/ml)
- Chloroform-isoamyl alcohol (24:1)
- 3M Sodium Acetate pH 5.2
- 70% Ethanol
- TE buffer
- Buffered Phenol
- 1.5 ml microcentrifuge tubes

Methods

Gloves were worn at all time to prevent contamination. The cotton swabs were cut and inserted into tubes labeled accordingly. Each tube containing the cut cotton was added with 500 µl of digestion buffer and 12 µl of proteinase-K and then incubated overnight at 56ºC. Another 12 µl of proteinase-K was added into each tube again after 2 hours of incubation. On the next day, 120 µl of phenol was added directly into each tube then mixed vigorously and spun at 10,000 rpm for 3 minutes. The aqueous phase was transferred to a new 1.5 ml tube using a cut-tip (pipette tip with the sharp end removed). After that, 250 µl of phenol and 250 µl of chloroform-isoamyl alcohol were

added into each tube, mixed vigorously and spun at 10,000 rpm for 3 minutes. The aqueous layer on top was transferred to a new 1.5 ml tube using another cut-tip. Each new tube was added with 250 µl of chloroform-isoamyl alcohol, mixed vigorously and spun at 10,000 rpm for 5 minutes. The aqueous phase was carefully transferred to a new 1.5 ml tube using cut-tip. Each new tube was added with 500 µl of chilled ethanol and 50 µl of 2 M sodium acetate. The solution in each tube was gently mixed, inverted and spun at 10,000 rpm for 10 minutes. The supernatant was discarded and 500 µl of 70% ethanol was added. The DNA pellet was dislodged and spun at 10000 rpm for 3 minutes. The supernatant was discarded and the pellet was left to dry at room temperature. Each DNA pellet was added with 50 µl of TE buffer and incubated overnight at 30ºC. Extracted DNA was stored at -20ºC.

3.3 PCR Amplification

Materials

- Thermal cycler / PCR machine (MJ Research, USA)
- Microcentrifuge (Eppendorf, USA)
- *Taq* DNA polymerase (1u/µl) (Promega, USA)
- Nuclease-Free water (E-Pure water)
- 0.2 ml or 0.5 ml microcentrifuge tubes (compatible with thermal cycler)
- 1.5 ml microcentrifuge tubes
- Aerosol-resistant tips
- Ice
- Multiplex PCR kits (Promega, USA):
    - CTT Multiplex (CSF1PO, TPOX, TH01)
    - FFv Multiplex (F13A01, FESFPS, vWA)
    - Silver STR III Multiplex (D16S539, D7S820, D13S317)

Methods

Gloves and aerosol-resistant pipette tips were used to prevent cross-contamination. STR 10X Buffer and STR 10X Primer Pairs were thawed and kept on ice. Clean autoclaved 1.5 ml tube for PCR master mix and 0.5 ml microcentrifuge tubes for each reaction was placed into a rack and labeled appropriately. The number of reaction to be set up was determined. One or 2 reactions were added to this number to compensate for pipetting error. The required amount of each component for the PCR master mix was calculated by multiplying the number of reaction needed, to the volume of each component in a single reaction, in order to obtain the master mix final volume. The final volume of each component was added to the sterile 1.5 ml tube. The solution

was gently mixed and placed on ice. Each 0.5 ml microcentrifuge tube was added with 11.25 µl of PCR master mix (11.00 µl for FFv Multiplex) and placed on ice. Each 0.5 ml microcentrifuge tube contained 7.25 µl of E-Pure water (7.00 µl for FFv), 1.25 µl of STR 10× Buffer, 1.25 µl of Multiplex 10× primer Pair Mix, 0.50 µl of *Taq* DNA polymerase, and 1.00 µl of $MgCl_2$ (2.0 mM).

DNA sample of 1.25 µl each (1.50 µl for FFv) was pipetted into the respective tube containing 11.25 µl of PCR master mix (11.00 µl for FFv). As for positive amplification control, 1.25 µl of K562 DNA (1.50 µl for FFv) was pipetted into a 0.5 ml microcentrifuge tube, containing 11.25 µl of PCR master mix (11.00 µl for FFv). As for negative amplification control, 1.25 µl of sterile water (1.50 µl for FFv) was pipette into a 0.5 ml microcentrifuge tube, containing 11.25 µl of PCR master mix (11.00 µl for FFv). PCR was run immediately after mixing of components.

For CTT Multiplex, PCR was started on initial denaturation at 96°C for 2 minutes followed by 10 cycles of denaturation at 64°C for 1 minute, annealing at 64°C for 1 minute, and extention at 70°C for 1.5 minutes. The next 20 cycles then begun at 90°C for 1 minute, 64°C for another minute, and then 70°C for 1.5 minutes. Finally the amplified products were kept at 4°C. As for FFv and STR III Multiplex, PCR was started on initial denaturation at 96°C for 2 minutes, followed by 10 cycles of denaturation at 94°C for 1 minute, annealing at 60°C for 1 minute, and extention at 70°C for 1.5 minutes. The next 20 cycles then begun at 90°C for 1 minute, then 60°C for another minute, and then 70°C for 1.5 minutes. The PCR products were then put on extension at 60°C for 30 minutes. Finally the amplified products were kept at 4°C.

3.4 Polyacrylamide Gel Preparation

Materials

(Solution compositions are in Appendix)

- 40% acrylamide:bisacrylamide (19:1) and TEMED
- 10× TBE buffer
- 10% Ammonium Persulfate
- Urea
- Bind silane
- 0.5% acetic acid in 95% ethanol
- Tissue culture filter (0.2 micron)
- Polyacrylamide gel electrophoresis apparatus for gels ≥30 cm
- Glass plates and side spacers for polyacrylamide gel ≥30 cm
- 14 cm vinyl doublefine sharkstooth comb, 30 point, 0.4 mm thick
- Power supply
- Detergent
- Clamps
- Diamond pencil for marking glass plates
- Water repellent solution
- 50 ml syringe and needle

Methods

Each glass plate was etched on one side at one corner with a diamond pencil to distinguish the treated sides of the glass plates. The shorter and longer glass plates were cleaned twice with 95% ethanol and tissues. Using gloves, 3 ml of water repellent solution was applied onto the etched side of the longer glass plate. The water repellent

was spread with a dry paper towel using a circular motion over the entire surface and left to dry for 5 minutes. The excess water repellent solution was removed with a paper towel saturated with deionized water. Finally, the glass plate was dried with tissue.

In a chemical fume hood, fresh binding solution was prepared by adding 3 µl of bind saline to 1 ml of 0.5% acetic acid in 95% ethanol in a 1.5 ml microcentrifuge tube. The etched side of the shorter glass plate was wiped entirely using a piece of tissue saturated with the freshly prepared binding solution and left to dry for 5 minutes. The shorter glass plate was wiped for 3 - 4 times with 95% ethanol and tissues to remove the excess binding solution. Precaution was taken as not to allow the treated surface of the both glass plates to touch each other. The glass plates were assembled by placing 0.4 mm side spacers between the plates. Assembled glass plates were sealed with sealing tapes at both sides and bottom to prevent leakage. Clamps were used to hold them in place. The assembled plates were leaned against a test tube rack.

In order to prepare 60 ml of 6% acrylamide solution, 25.2 g of Urea, 29.0 ml of dH_2O, 3 ml of 10× TBE Buffer and 9 ml of 40% Acrylamide:Bisacrylamide (19:1) were mixed together. The acrylamide solution was filtered through a 0.2 micron filter before 50 µl of TEMED and 500 µl of 20% ammonium persulfate were added to the acrylamide solution and mixed gently. Immediately, the treated acrylamide solution was sucked into a syringe. A needle was put on the filled syringe and the acrylamide solution was injected carefully between the glass plates, starting at one side of the assembled plates at a constant flow. The gel was then placed horizontally on two test tube racks. Any bubble that formed was removed out. One sharkstooth comb, straight side (without comb teeth), was inserted into the gel between the glass plates. The comb was secured with clamps. The remaining acrylamide solution was drained into a

disposable conical tube as a polymerization control. The syringe and the needle were rinsed thoroughly with water.

The plates were left for at least 1 hour for the acrylamide solution to polymerize. The polymerization control was checked to be sure that polymerization had occurred. When the solution was no longer watery and attached firmly intact, the polymerized control was disposed and the plates were left for another 1 hour for the acrylamide solution to fully polymerize.

3.5 Polyacrylamide Gel Electrophoresis

Methods

Gel Pre-run

The clamps were removed from the polymerized acrylamide gel and the glass plates were cleaned with paper towels saturated with deionized water. Any excess polyacrylamide was shaved away from the comb. The comb was removed. 0.5× TBE was added to the bottom chamber of the electrophoresis apparatus. The gel (glass plates) was gently lowered into the buffer with the longer plate facing out and the well-sided on top. The glass plates were secured to the sequencing gel apparatus. 0.5× TBE was added to the top buffer chamber of the electrophoresis apparatus, covering the wells on top of the gel. Air bubbles and small pieces of polyacrylamide were removed on the top of the gel and also the well area. The gel was pre-run at 60 – 65 watts for 20 – 30 minutes.

Sample Preparation

The amplified PCR samples were prepared by mixing 2.5 µl of each sample with 2.5 µl of STR 2× Loading Solution in each microcentrifuge tube labeled accordingly. Number of sample groups for positive control, negative control and STR Ladder were determined. This number was then separately multiplied by 2.5 µl for each of the positive control, negative control and STR Ladder group respectively; and also by 2.5 µl of STR 2× Loading Solution, in order to obtain the total loading volume for each group. The total loading volumes for each group were prepared in microcentrifuge tubes labeled accordingly. All the samples were spun briefly in a microcentrifuge to pellet the contents.

Sample Loading

The samples were denatured by heating at 95°C for 2 minutes, then immediately chilled on ice. After the pre-run, the well area of the gel was flushed using a 100cc syringe. One to 2 mm of the sharkstooth comb was inserted carefully into the gel. The comb was left inserted in the gel during both gel loading and electrophoresis. Three or 4 µl of each sample was loaded into the respective wells. The loading process was done no longer than 20 minutes to prevent the gel from cooling.

Gel Electrophoresis

At the completion of loading, the electrophoresis was run using the same conditions as in the pre-run, which is 60 – 65 watts. In a 6% gel, bromophenol blue migrates at approximately 25 bases and xylene cyanol migrates at approximately 105 bases. Knowing the size ranges for each locus (see Appendix) and migration characteristics of the dyes, electrophoresis was stopped any time after the locus of interest has passed the midpoint of the gel.

Reuse of Glass Plates

The plates were immersed in a 10% NaOH solution for at least 1 hour. The gel was discarded and the glass plate was cleaned with deionized water and a detergent. The 10 % NaOH solution was reused for additional gels. All cleaning utensils for the longer glass plates were kept separate from those for the shorter glass plates to prevent cross-contamination of the binding solution. The shorter glass plate preparation was repeated for each gel. The longer glass plate preparation was repeated after every four gels.

3.6 Silver Staining

Materials

Fix/stop solution

- Staining solution
- Developer solution
- Wash tubs (appropriate size according to system requirements)

Notes:

Steps involving formaldehyde were performed in a chemical hood.

Methods

In order to prepare 600 ml of Fix/Stop solution, 120 ml of glacial acetic acid and 480 ml of dH_2O were mixed together. As for 600 ml of Staining solution, 0.6 g of silver nitrate ($AgNO_3$) was dissolved in 400 ml of dH_2O and 0.9 ml of formaldehyde in a dark bottle. Finally, dH_2O was added to make up to 600 ml and kept at 4°C. Dark bottle was also used to prepare 600 ml of Developer solution by dissolving 18 g of sodium carbonate (Na_2CO_3) in 400 ml of dH_2O, then adding 0.6 ml of formaldehyde and 120 µl of sodium thiosulfate ($Na_2S_2O_3 \cdot 5H_2O$). Finally, dH_2O was added to make up to 600 ml and kept at 4°C.

After electrophoresis, the glass plates were removed from the apparatus and placed on a flat surface. The comb and the side spacers were removed. The two glass plates were separated carefully using a plastic wedge. The gel (attached to the shorter plate) was placed in a shallow plastic tray and silver stained.

The gel was placed in a tray filled with Fix/Stop solution and left to soak for 20 minutes. The plastic tray was gently rocked from time to time. The gel was then transferred to another plastic tray filled with deionized water and rocked gently for 2 minutes. Soaking the gel in deionized water was repeated twice with new fresh water used each time. The gel was then transferred to another tray filled with Staining solution and left to soak for 30 minutes. The tray was covered and gently rocked from time to time. The gel was then briefly soaked in deionized water for about 3 seconds and transferred to another tray filled with Developer solution. The tray was covered and gently rocked from time to time. The gel was left soaked in Developer solution for 10 minutes or until alleles and ladders were visible. When the alleles and ladders were seen, the gel was transferred back to the tray filled with Fix/Stop solution and left to soak for 5 minutes. The gel was then briefly soaked in deionized water for 2 minutes and placed upright to dry overnight at room temperature. The gel was ready to be analyzed the next day when it dried. Genotype observed on the gel at each locus for each sample was recorded accordingly.

4.0 RESULTS AND DATA ANALYSIS

4.1 Results from Polyacrylamide Gel Electrophoresis (PAGE)

The genotypes obtained from the three multiplex PCR (CTT, FFv, and SilverSTR III) for the 109 Sikh individuals are recorded in Table 4.1.1, Table 4.1.2 and Table 4.1.3, and the genotype frequencies for the STR are in Table 4.3. The genotype frequencies were estimated by the method of maximum likelihood, whereby gene frequencies for two allelic (P1 and P2) co-dominant system is:

$$P1 = (2x + y/2N) \text{ and}$$

$$P2 = 1 - P1$$

Where 'x' is the number of homologous type of one allele; 'y' is the number of heterozygous type and 'N' is the total number of individuals analyzed. Gene frequencies were computed and Chi-square test was performed to assess Hardy-Weinberg equilibrium of the population studied. The raw data was analyzed using computer software DNA-View (Brenner, 2002).

Figure 4.1a, b, and c show the example of gels after running PAGE for CTT, FFv and STR III multiplex PCR amplification products, respectively. The complete genotype result of all 109 Sikh individuals for CTT, FFv and STR III multiplex is recorded in Table 4.1.

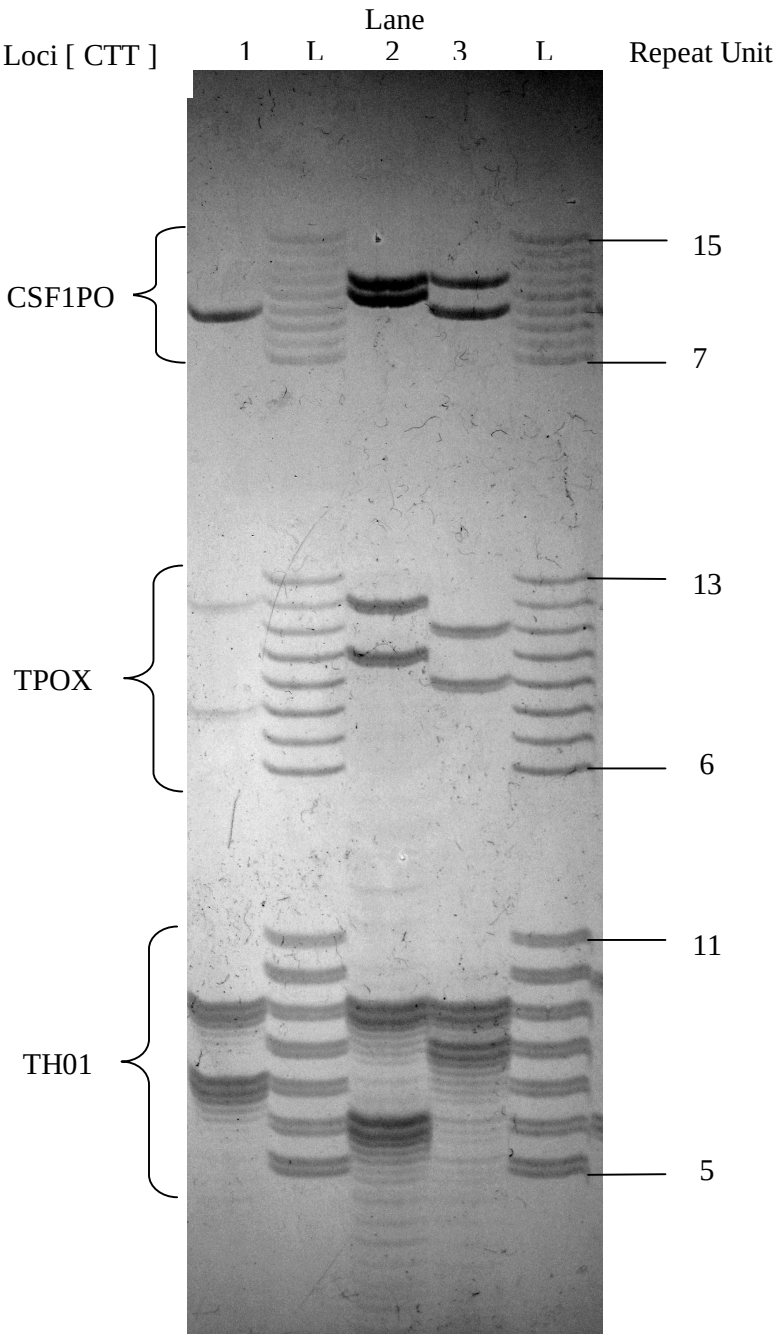

Figure 4.1a: PAGE of CTT multiplex PCR amplification products

Lane 1 : DNA genotypes CSF1PO 10,10; TPOX 8,12; TH01 7,9.
Lane 2 : DNA genotypes CSF1PO 11,12; TPOX 10,12; TH01 6,9.
Lane 3 : DNA genotypes CSF1PO 10,12; TPOX 9,11; TH01 8,9.
Lane L : STR allelic ladders for CTT multiplex.
Locus CSF1PO usually contains alleles of 7 – 15 repeats;
Locus TPOX usually contains alleles of 6 – 13 repeats;
Locus TH01 usually contains alleles of 5 – 11 repeats.

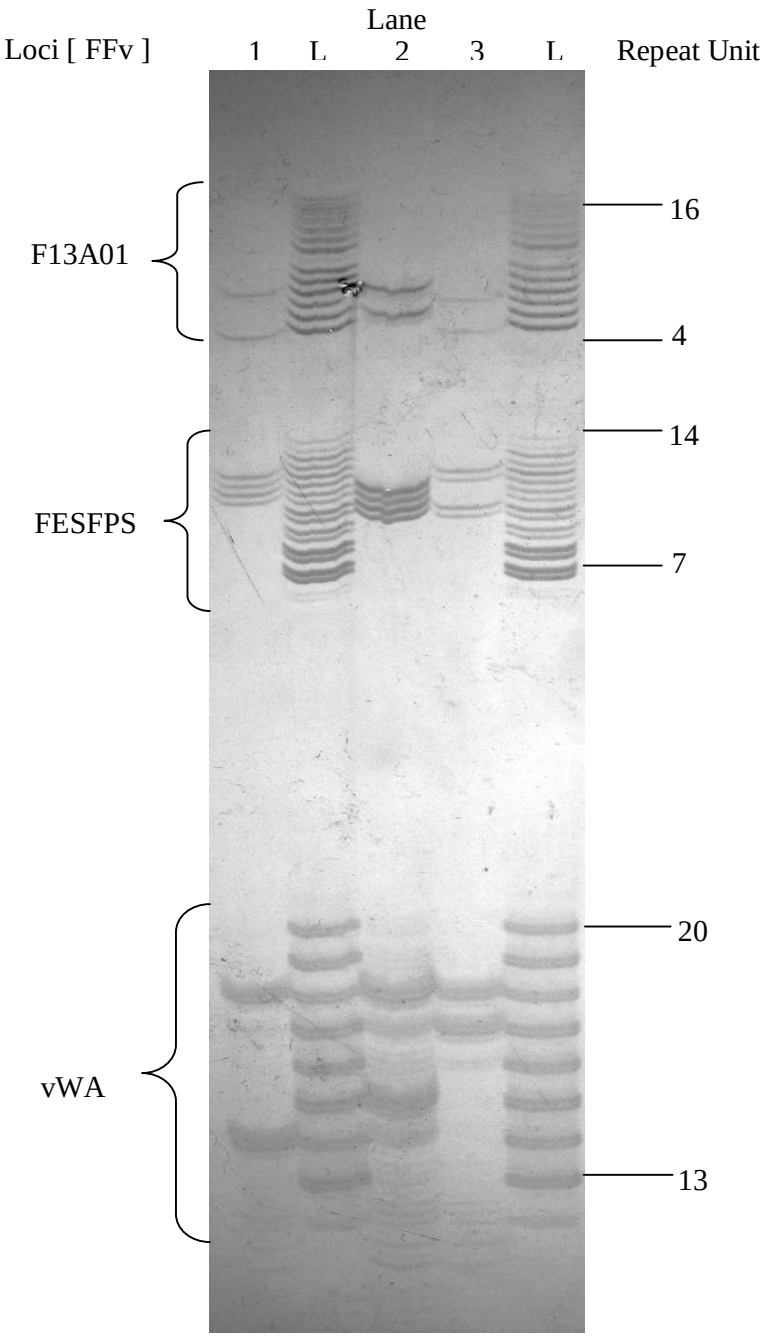

Figure 4.1b: PAGE of FFv multiplex PCR amplification products

Lane 1 : DNA genotypes F13A01 3.2,7; FESFPS 11,12; vWA 14,18.
Lane 2 : DNA genotypes F13A01 5,7; FESFPS 10,11; vWA 15,18.
Lane 3 : DNA genotypes F13A01 4,6; FESFPS 10,12; vWA 17,18.
Lane L : STR allelic ladders for FFv multiplex.
Locus F13A01 usually contains alleles of 4 – 16 repeats;
Locus FESFPS usually contains alleles of 7 – 14 repeats;
Locus vWA usually contains alleles of 13 – 20 repeats.

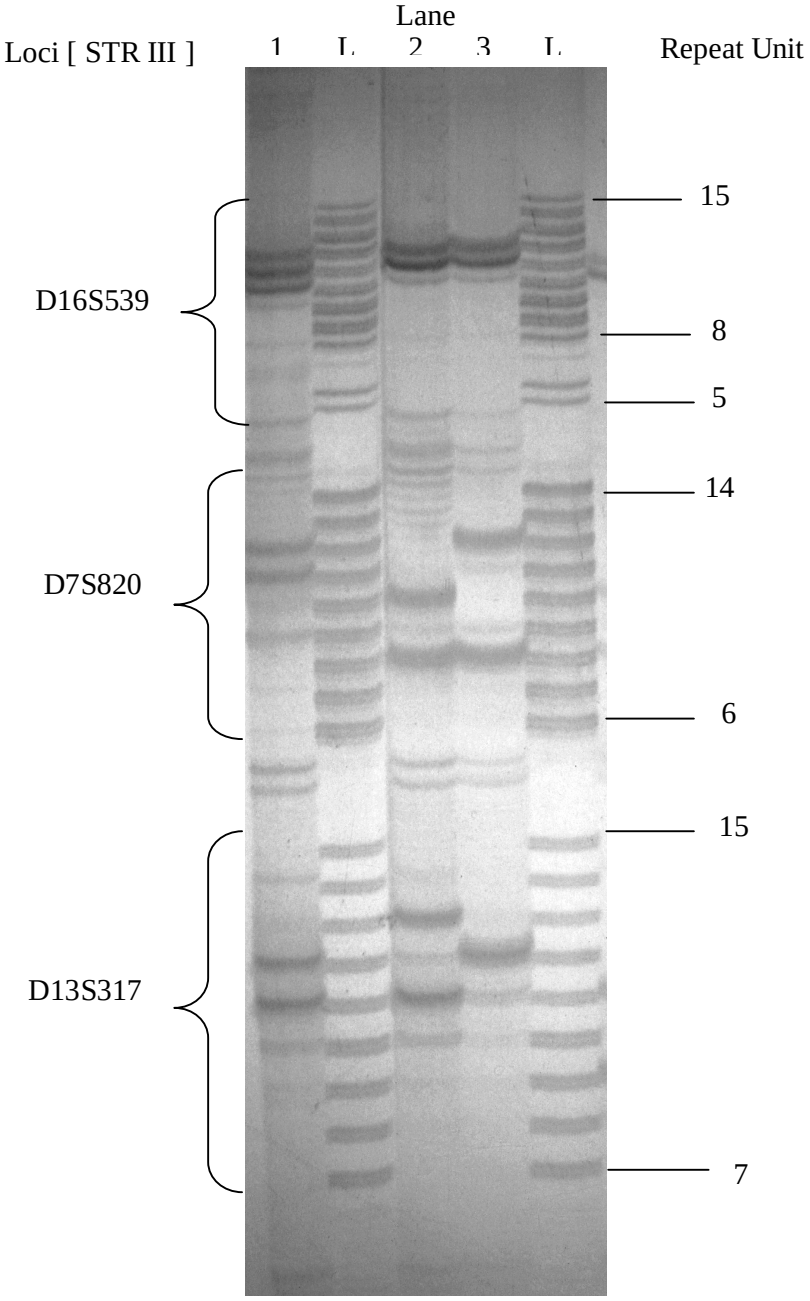

Figure 4.1c: PAGE of STR III multiplex PCR amplification products

Lane 1 : DNA genotypes D16S539 11,12; D7S820 11,12; D13S317 11,12.
Lane 2 : DNA genotypes D16S539 12,12; D7S820 8,10; D13S317 11,13.
Lane 3 : DNA genotypes D16S539 12,12; D7S820 8,12; D13S317 12,12.
Lane L : STR allelic ladders for STR III multiplex.
Locus D16S539 usually contains alleles of 5, 8 – 15 repeats;
Locus D7S820 usually contains alleles of 6 – 14 repeats;
Locus D13S317 usually contains alleles of 7 – 15 repeats.

Table 4.1: Genotype Observed for Individual Sample

Sample No	CTT Multiplex			FFv Multiplex			SilverSTR III Multiplex		
	CSF1PO	TPOX	TH01	F13A01	FESFPS	vWA	D16S539	D7S820	D13S317
1	11,12	8,9	7,7	6,8	9,11	15,16	12,13	8,11	13,13
2	10,12	8,8	8,9	7,7	10,12	17,17	11,11	9,10	12,13
3	10,10	8,12	7,9	3.2,7	11,12	14,18	11,11	11,12	11,12
4	11,11	8,10	6,6	3.2,5	11,12	17,18	9,12	10,11	8,10
5	11,11	8,11	6,8	5,6	11,12	16,16	10,10	8,11	8,13
6	12,12	8,9	6,9	5,5	11,12	14,18	8,10	9,10	8,11
7	11,12	9,9	9,10	4,6	12,12	17,19	10,10	11,12	11,12
8	12,12	9,9	7,7	4,5	12,13	16,18	9,10	10,12	11,11
9	10,11	8,11	6,9	3.2,6	10,10	17,18	13,13	8,9	12,12
10	10,10	11,11	6,8	5,5	8,12	14,17	8,12	8,12	11,12
11	12,13	9,11	6,9.3	5,7	10,12	14,15	8,11	10,11	9,11
12	11,11	8,10	9,9.3	3.2,5	10,12	16,17	10,12	9,13	8,11
13	10,12	8,8	6,7	4,4	10,11	16,17	11,14	8,10	12,12
14	10,11	8,9	8,9	6,7	11,12	16,17	9,10	11,11	11,11
15	12,12	8,10	6,6	4,5	10,10	17,19	11,13	8,10	12,13
16	10,12	8,12	8,10	5,6	11,12	17,20	10,10	10,12	10,11
17	11,12	8,11	6,9	5,5	11,12	14,16	9,13	8,10	9,12
18	9,10	8,11	8,9	3.2,6	10,12	16,19	9,13	11,12	8,11
19	11,12	8,11	7,9.3	5,5	11,12	18,18	9,12	8,12	8,13
20	11,12	8,11	8,8	4,5	11,12	17,19	12,13	7,12	8,12
21	12,13	8,9	6,7	3.2,15	10,10	17,17	9,12	11,11	11,11
22	10,12	8,11	6,9	5,7	11,12	15,17	11,12	10,12	9,12
23	10,13	8,8	9,9.3	7,7	11,12	18,19	9,11	10,12	8,8
24	11,12	8,9	7,10	5,7	10,13	16,17	9,11	9,11	9,11
25	11,12	10,12	6,9	5,7	10,11	15,18	12,12	8,10	11,13
26	11,12	8,9	6,6	3.2,7	10,10	18,18	11,11	8,10	8,9
27	10,11	8,10	8,9	3.2,5	121,12	17,17	8,11	10,11	11,11
28	12,12	8,11	6,7	4,7	10,12	18,18	9,11	8,11	9,12
29	12,12	11,11	8,9.3	5,7	11,11	17,18	12,13	8,9	8,11
30	10,12	9,11	9.3,9.3	3.2,6	11,12	14,17	10,12	7,11	8,13
31	11,11	8,10	9,9.3	5,6	11,13	16,17	10,12	9,10	12,12
32	10,12	7,11	8,8	3.2,5	11,11	15,18	11,13	8,10	9,12
33	12,13	8,8	9,9.3	4,7	10,11	16,18	11,12	6,8	11,11
34	10,12	8,11	8,9	4,6	11,11	16,18	12,13	8,10	8,9
35	11,13	9,11	9,9.3	5,5	11,12	14,17	11,11	8,12	8,12
36	12,12	8,9	6,7	5,7	10,11	17,18	11,11	11,12	8,11
37	12,12	8,10	8,9	5,5	10,12	14,17	10,11	9,12	13,13
38	10,12	8,10	9.3,9.3	3.2,5	10,12	13,17	8,8	10,11	10,13
39	10,11	8,11	9,9	3.2,5	11,11	16,18	9,11	12,13	11,12
40	11,12	8,10	9,9	3.2,6	12,12	16,17	11,12	8,10	8,12
41	10,11	8,8	6,7	5,5	10,12	14,17	11,13	10,12	11,12
42	11,12	9,11	6,9.3	4,5	10,12	15,17	11,11	9,11	11,14
43	10,11	9,11	6,9.3	4,6	12,12	14,17	11,13	9,11	10,11
44	11,13	9,11	6,8	3.2,6	10,10	16,17	9,11	11,12	11,12
45	11,13	9,9	9,9.3	5,6	7,11	13,16	11,12	11,12	12,12
46	10,12	8,11	6,7	3.2,5	10,11	14,17	11,12	10,12	8,12
47	12,12	8,8	8,9.3	5,16	10,11	17,17	11,11	8,11	8,12
48	11,13	8,11	6,7	6,6	11,11	17,17	13,13	9,11	10,11
49	11,12	8,9	7,9.3	4,6	11,12	16,17	9,11	8,10	12,13
50	12,15	8,11	7,9.3	5,6	10,12	16,16	12,14	10,12	11,12
51	10,12	8,11	6,9.3	6,6	12,12	16,17	9,12	8,12	9,9
52	10,11	11,11	6,9	4,4	11,11	16,18	9,12	8,11	8,12
53	11,12	8,8	7,9	3.2,5	11,12	15,17	10,12	8,11	8,11

54	11,12	8,11	6,9.3	5,5	11,12	17,18	10,11	9,10	8,12
55	12,12	11,12	6,9	5,5	11,12	14,14	9,11	8,10	11,11
Sample	CTT Multiplex			FFv Multiplex			SilverSTR III Multiplex		
No	CSF1PO	TPOX	TH01	F13A01	FESFPS	vWA	D16S539	D7S820	D13S317
56	10,11	8,10	7,9	3.2,5	11,13	16,16	9,12	8,10	11,11
57	12,12	11,11	6,6	3.2,5	11,12	14,18	9,13	8,12	12,14
58	12,12	12,12	6,6	5,5	11,12	17,17	11,12	9,12	11,11
59	11,12	10,11	6,7	5,6	11,12	14,17	12,13	7,8	12,13
60	10,13	10,11	8,9	4,16	11,13	14,16	10,11	11,12	11,11
61	12,12	8,8	6,10	5,5	11,11	14,18	9,11	9,10	11,12
62	10,12	9,11	8,9	3.2,6	10,12	17,18	11,11	8,12	12,15
63	9,12	8,9	6,8	3.2,16	11,11	15,19	9,11	8,11	11,11
64	9,11	9,11	6,9	3.2,6	10,11	16,16	10,11	10,11	8,12
65	12,12	8,11	6,6	7,7	11,14	15,16	11,12	8,9	12,15
66	10,12	7,11	8,9	15,16	10,10	15,18	11,12	9,9	12,13
67	11,11	9,12	9,9	3.2,16	11,11	18,19	11,13	8,10	12,12
68	10,11	8,10	7,8	3.2,6	10,10	16,17	9,10	8,8	8,11
69	10,12	10,10	9,9.3	6,7	11,11	16,17	11,11	10,10	11,12
70	12,12	10,10	9.3,9.3	5,5	10,11	17,20	9,11	8,10	10,11
71	12,12	8,8	6,7	5,6	10,13	16,17	12,13	10,11	8,8
72	10,12	8,10	7,10	3.2,7	11,12	14,16	9,11	8,11	12,14
73	10,13	8,9	6,6	3.2,6	12,12	16,19	12,12	10,13	9,12
74	11,12	8,11	6,10	3.2,14	10,11	15,17	11,11	9,11	8,12
75	11,12	8,9	6,7	5,7	11,13	16,18	11,14	8,10	8,13
76	12,12	8,11	6,10	5,5	11,11	14,17	13,14	9,9	8,11
77	10,10	8,8	6,10	5,5	10,11	16,17	11,11	8,10	9,11
78	12,14	8,8	6,7	5,7	11,11	14,15	8,12	8,12	10,11
79	10,12	8,11	6,9	7,7	11,14	15,16	11,12	9,10	12,12
80	10,12	8,8	9,9	7,16	10,11	14,15	10,14	11,11	9,11
81	10,12	8,11	6,9	7,7	12,14	15,16	11,11	8,9	12,12
82	12,12	8,10	9,10	5,6	11,11	15,16	10,11	8,10	10,12
83	11,12	7,8	8,9	7,15	10,10	16,18	8,12	8,9	8,13
84	11,12	8,11	9,9.3	4,5	10,13	15,16	10,12	8,10	8,13
85	12,13	8,11	7,8	4,7	10,12	17,18	10,11	9,11	9,13
86	12,13	8,11	8,9.3	7,16	10,11	17,17	8,10	8,12	11,12
87	10,13	8,11	6,9.3	5,6	11,11	17,19	13,13	11,12	11,12
88	11,13	8,11	8,9	5,6	11,12	17,18	11,12	10,10	8,8
89	11,12	8,11	8,9.3	5,6	12,12	18,18	10,10	9,11	11,13
90	10,11	8,10	6,9.3	6,7	10,11	16,17	12,12	10,12	11,12
91	11,12	9,9	7,8	6,7	11,13	14,17	9,12	8,11	9,9
92	11,12	8,11	7,9.3	5,6	10,12	16,17	12,12	8,8	11,11
93	11,12	9,11	9,9.3	4,7	10,11	14,17	8,11	10,13	8,11
94	10,13	8,8	7,9	6,6	11,12	15,18	12,12	10,12	12,13
95	10,11	8,11	6,9.3	6,6	10,12	17,18	8,11	8,10	8,11
96	10,12	9,11	9.3,9.3	5,16	8,10	17,18	12,14	8,9	8,12
97	10,12	8,9	8,9	5,5	11,11	14,19	12,12	10,12	8,14
98	10,11	8,10	9,9.3	5,7	11,13	16,17	10,13	8,10	11,12
99	11,13	8,9	6,9	4,7	11,11	14,16	9,9	9,12	11,12
100	11,12	8,8	9,9.3	5,6	10,10	15,17	9,11	8,12	11,13
101	11,11	8,11	9.3,9.3	5,7	10,11	16,16	9,9	11,12	10,12
102	12,12	8,11	6,9.3	5,5	12,12	17,18	10,13	10,13	8,12
103	10,12	8,8	7,9.3	5,6	10,12	18,18	9,11	8,11	9,11
104	11,12	8,10	6,9	6,6	10,10	12,17	9,9	8,12	8,12
105	10,11	8,8	7,9.3	4,7	10,11	15,16	8,9	12,12	12,12
106	11,12	8,11	6,7	3.2,3.2	11,11	18,18	11,12	8,8	8,12
107	11,11	8,10	9,9	5,7	10,12	15,17	10,13	8,10	12,13
108	11,11	11,12	8,9.3	5,6	11,12	17,18	10,12	8,10	10,11
109	11,11	8,10	6,6	3.2,5	10,11	14,16	11,12	11,12	12,12

4.1.1 Genotype Frequencies from CTT Multiplex PCR

Table 4.1.1a: CSF1PO

Genotype	Observed numbers
9 – 10	1
9 – 11	1
9 – 12	1
10 – 10	3
10 – 11	13
10 – 12	20
10 – 13	5
11 – 11	9
11 – 12	25
11 – 13	6
12 – 12	18
12 - 13	5
12 – 14	1
12 – 15	1
Total	109

Table 4.1.1b: TPOX

Genotype	Observed numbers
7 – 8	1
7 – 11	2
8 – 8	16
8 – 9	13
8 – 10	17
8 – 11	31
8 – 12	2
9 – 9	4
9 – 11	10
9 – 12	1
10 – 10	2
10 – 11	2
10 – 12	1
11 – 11	4
11 – 12	2
12 – 12	1
Total	109

Table 4.1.1c: THO1

Genotype	Observed numbers
6 – 6	8
6 – 7	12
6 – 8	4
6 – 9	12
6 – 9.3	9
6 – 10	4
7 – 7	2
7 – 8	3
7 – 9	4
7 – 9.3	6
7 – 10	2
8 – 8	2
8 – 9	12
8 – 9.3	5
8 – 10	1
9 – 9	5
9 – 9.3	11
9 – 10	2
9.3 – 9.3	5
Total	109

4.1.2 Genotype Frequencies from FFv Multiplex PCR

Table 4.1.2a: F13A01

Genotype	Observed numbers
3.2 – 3.2	1
3.2 – 5	11
3.2 – 6	9
3.2 – 7	3
3.2 – 14	1
3.2 – 15	1
3.2 – 16	2
4 – 4	2
4 – 5	5
4 – 6	4
4 – 7	6
4 – 16	1
5 – 5	16
5 – 6	15
5 – 7	11
5 – 16	2
6 – 6	5
6 – 7	4
6 – 8	1
7 – 7	5
7 – 15	1
7 – 16	2
15 – 16	1
Total	109

Table 4.1.2b: FESFPS

Genotype	Observed numbers
7 – 11	1
8 – 10	1
8 – 12	1
9 – 11	1
10 – 10	10
10 – 11	17
10 – 12	16
10 – 13	3
11 – 11	17
11 – 12	24
11 – 13	6
11 – 14	2
12 – 12	8
12 – 13	1
12 – 14	1
Total	109

Table 4.1.2c: vWA

Genotype	Observed numbers
12 – 17	1
13 – 16	1
13 – 17	1
14 – 14	1
14 – 15	3
14 – 16	5
14 – 17	11
14 – 18	4
14 – 19	1
15 – 16	7
15 – 17	6
15 – 18	4
15 – 19	1
16 – 16	5
16 – 17	16
16 – 18	7
16 – 19	2
17 – 17	7
17 – 18	12
17 – 19	4
17 – 20	2
18 – 18	6
18 – 19	2
Total	109

4.1.3 Genotype Frequencies from SilverSTR III Multiplex PCR

Table 4.1.3a: D16S539

Genotype	Observed numbers
8 – 8	1
8 – 9	1
8 – 10	2
8 – 11	4
8 – 12	3
9 – 9	3
9 – 10	3
9 – 11	13
9 – 12	7
9 – 13	3
10 – 10	4
10 – 11	6
10 – 12	6
10 – 13	3
10 – 14	1
11 – 11	12
11 – 12	12
11 – 13	5
11 – 14	2
12 – 12	6
12 – 13	6
12 – 14	2
13 – 13	3
13 – 14	1
Total	109

Table 4.1.3b: D7S820

Genotype	Observed numbers
6 – 8	1
7 – 8	1
7 – 11	1
7 – 12	1
8 – 8	3
8 – 9	6
8 – 10	21
8 – 11	10
8 – 12	10
9 – 9	2
9 – 10	6
9 – 11	7
9 – 12	3
9 – 13	1
10 – 10	2
10 – 11	6
10 – 12	10
10 – 13	3
11 – 11	3
11 – 12	10
12 – 12	1
12 – 13	1
Total	109

Table 4.1.3c: D13S317

Genotype	Observed numbers
8 – 8	3
8 – 9	2
8 – 10	1
8 – 11	10
8 – 12	13
8 – 13	6
8 – 14	1
9 – 9	2
9 – 11	5
9 – 12	5
9 – 13	1
10 – 11	6
10 – 12	2
10 – 13	1
11 – 11	11
11 – 12	14
11 – 13	3
11 – 14	1
12 – 12	9
12 – 13	7
12 – 14	2
12 – 15	2
13 – 13	2
Total	109

4.2 Processing the Results

Results of genotype for all the samples are processed to calculate allele frequencies and expected numbers of every locus. As an example, the result for STR locus CSF1PO in Table 4.1.1a is processed as below:

Allele	Number of times appeared	Total number of times appeared, *x*	Allele frequency, *x*/218
9	1+1+1	3	0.0138
10	1+2(3)+13+20+5	45	0.2064
11	1+13+2(9)+25+6	63	0.2890
12	1+20+25+2(18)+5+1+1	89	0.4083
13	5+6+5	16	0.0734
14	1	1	0.0046
15	1	1	0.0046
		$\sum = 218$	$\sum = 1.0001$

Based on Table 4.2.1a, in a sample of 109 Sikhs, allele 9 appeared once in each of the genotype 9–10, 9–11 and 9–12. Hence, the number of times allele 9 appeared is 1+1+1=3. As for allele 10, the number of times appeared is 1+2(3)+13+20+5=45. 2(3) being obtained because genotype 10–10 is homozygous for allele 10, thus 2×3. Total number of alleles in 109 Sikhs is 218, twice the number of individuals analyzed because every person has a pair of alleles for each locus. The total number of times each allele appeared, *x*, is then divided by 218 to obtain the respective allele frequency. These values are then used to calculate the expected number, E, of each genotype, according to the Hardy-Weinberg Law. For heterozygotes, such as genotype 9–10 at CSF1PO:

Frequency of $A_1A_2 = 2 \cdot p_1 \cdot p_2$ p_1, probability for allele 9 = 0.0138

p_2, probability for allele 10 = 0.2064

Frequency of genotype 9–10 = 2 (0.0138) (0.2064)

E of genotype 9–10 = 2 (0.0138) (0.2064) (number of individuals involved)

= 2 (0.0138) (0.2064) (109)

= 0.6209

For homozygotes, such as genotype 10–10:

Frequency of $A_1A_1 = p_1^2$ p_1, probability for allele 10 = 0.2064

Frequency of genotype 10–10 = $(0.2064)^2$

E of genotype 10–10 = $(0.2064)^2$ (number of individuals involved)

= $(0.2064)^2$ (109)

= 4.6435

The same method is used to calculate the expected numbers for other genotypes and loci. The results are shown below.

Table 4.2.1a: Allele frequency at locus CSF1PO

Allele	Number of times appeared	Total number of times appeared, *x*	Allele frequency, *x*/218
9	1+1+1	3	0.0138
10	1+2(3)+13+20+5	45	0.2064
11	1+13+2(9)+25+6	63	0.2890
12	1+20+25+2(18)+5+1+1	89	0.4083
13	5+6+5	16	0.0734
14	1	1	0.0046
15	1	1	0.0046
		∑ = 218	∑ = 1.0001

Table 4.2.1b: Expected number for each genotype at locus CSF1PO

Genotype	Observed number, O	Expected number, E
9 – 10	1	0.62
9 – 11	1	0.87
9 – 12	1	1.23
10 – 10	3	4.64
10 – 11	13	13.00
10 – 12	20	18.37
10 – 13	5	3.30
11 – 11	9	9.10
11 – 12	25	25.72
11 – 13	6	4.62
12 – 12	18	18.17
12 – 13	5	6.53
12 – 14	1	0.41
12 – 15	1	0.41
Total	109	106.99

Table 4.2.2a: Allele frequency at locus TPOX

Allele	Number of times appeared	Total number of times appeared, *x*	Allele frequency, *x*/218
7	1+2	3	0.0138
8	1+2(16)+13+17+31+2	96	0.4404
9	13+2(4)+10+1	32	0.1468
10	17+2(2)+2+1	24	0.1101
11	2+31+10+2+2(4)+2	55	0.2523
12	2+1+1+2+2(1)	8	0.0367
		$\sum = 218$	$\sum = 1.0001$

Table 4.2.2b: Expected number for each genotype at locus TPOX

Genotype	Observed number, O	Expected number, E
7 – 8	1	1.32
7 – 11	2	0.76
8 – 8	16	21.14
8 – 9	13	14.09
8 – 10	17	10.57
8 – 11	31	24.22
8 – 12	2	3.52
9 – 9	4	2.35
9 – 11	10	8.07
9 – 12	1	1.17
10 – 10	2	1.32
10 – 11	2	6.06
10 – 12	1	0.88
11 – 11	4	6.94
11 – 12	2	2.02
12 – 12	1	0.15
Total	109	104.58

Table 4.2.3a: Allele frequency at locus TH01

Allele	Number of times appeared	Total number of times appeared, *x*	Allele frequency, *x*/218
6	2(8)+12+4+12+9+4	57	0.2615
7	12+2(2)+3+4+6+2	31	0.1422
8	4+3+2(2)+12+5+1	29	0.1330
9	12+4+12+2(5)+11+2	51	0.2339
9.3	9+6+5+11+2(5)	41	0.1881
10	4+2+1+2	9	0.0413
		$\sum = 218$	$\sum = 1.0000$

Table 4.2.3b: Expected number for each genotype at locus TH01

Genotype	Observed number, O	Expected number, E
6 – 6	8	7.45
6 – 7	12	8.11
6 – 8	4	7.58
6 – 9	12	13.33
6 – 9.3	9	10.72
6 – 10	4	2.35
7 – 7	2	2.20
7 – 8	3	4.12
7 – 9	4	7.25
7 – 9.3	6	5.83
7 – 10	2	1.28
8 – 8	2	1.93
8 – 9	12	6.78
8 – 9.3	5	5.45
8 – 10	1	1.20
9 – 9	5	5.96
9 – 9.3	11	9.59
9 – 10	2	2.11
9.3 – 9.3	5	3.86
Total	109	107.10

Table 4.2.4a: Allele frequency at locus F13A01

Allele	Number of times appeared	Total number of times appeared, *x*	Allele frequency, *x*/218
3.2	2(1)+11+9+3+1+1+2	29	0.1330
4	2(2)+5+4+6+1	20	0.0917
5	11+5+2(16)+15+11+2	76	0.3486
6	9+4+15+2(5)+4+1	43	0.1972
7	3+6+11+4+2(5)+1+2	37	0.1697
8	1	1	0.0046
14	1	1	0.0046
15	1+1+1	3	0.0138
16	2+1+2+2+1	8	0.0367
		$\sum = 218$	$\sum = 0.9999$

Table 4.2.4b: Expected number for each genotype at locus F13A01

Genotype	Observed number, O	Expected number, E
3.2 – 3.2	1	1.93
3.2 – 5	11	10.11
3.2 – 6	9	5.72
3.2 – 7	3	4.92
3.2 – 14	1	0.13
3.2 – 15	1	0.40
3.2 – 16	2	1.06
4 – 4	2	0.92
4 – 5	5	6.97
4 – 6	4	3.94
4 – 7	6	3.39
4 – 16	1	0.73
5 – 5	16	13.25
5 – 6	15	14.99
5 – 7	11	12.90
5 – 16	2	2.79
6 – 6	5	4.24
6 – 7	4	7.30
6 – 8	1	0.20
7 – 7	5	3.14
7 – 15	1	0.51
7 – 16	2	1.36
15 – 16	1	0.11
Total	109	101.01

Table 4.2.5a: Allele frequency at locus FESFPS

Allele	Number of times appeared	Total number of times appeared, *x*	Allele frequency, *x*/218
7	1	1	0.0046
8	1+1	2	0.0092
9	1	1	0.0046
10	1+2(10)+17+16+3	57	0.2615
11	1+1+17+2(17)+24+6+2	85	0.3899
12	1+16+24+2(8)+1+1	59	0.2706
13	3+6+1	10	0.0459
14	2+1	3	0.0138
		$\sum = 218$	$\sum = 1.0001$

Table 4.2.5b: Expected number for each genotype at locus FESFPS

Genotype	Observed number, O	Expected number, E
7 – 11	1	0.39
8 – 10	1	0.52
8 – 12	1	0.54
9 – 11	1	0.39
10 – 10	10	7.45
10 – 11	17	22.23
10 – 12	16	15.43
10 – 13	3	2.62
11 – 11	17	16.57
11 – 12	24	23.00
11 – 13	6	3.90
11 – 14	2	1.17
12 – 12	8	7.98
12 – 13	1	2.71
12 – 14	1	0.81
Total	109	105.71

Table 4.2.6a: Allele frequency at locus vWA

Allele	Number of times appeared	Total number of times appeared, *x*	Allele frequency, *x*/218
12	1	1	0.0046
13	1+1	2	0.0092
14	2(1)+3+5+11+4+1	26	0.1193
15	3+7+6+4+1	21	0.0963
16	1+5+7+2(5)+16+7+2	48	0.2202
17	1+1+11+6+16+2(7)+12+4+2	67	0.3073
18	4+4+7+12+2(6)+2	41	0.1881
19	1+1+2+4+2	10	0.0459
20	2	2	0.0092
		$\sum = 218$	$\sum = 1.0001$

Table 4.2.6b: Expected number for each genotype at locus vWA

Genotype	Observed number, O	Expected number, E
12 – 17	1	0.31
13 – 16	1	0.44
13 – 17	1	0.62
14 – 14	1	1.55
14 – 15	3	2.50
14 – 16	5	5.73
14 – 17	11	7.99
14 – 18	4	4.89
14 – 19	1	1.19
15 – 16	7	4.62
15 – 17	6	6.45
15 – 18	4	3.95
15 – 19	1	0.96
16 – 16	5	5.29
16 – 17	16	14.75
16 – 18	7	9.03
16 – 19	2	2.20
17 – 17	7	10.29
17 – 18	12	12.60
17 – 19	4	3.07
17 – 20	2	0.62
18 – 18	6	3.86
18 – 19	2	1.88
Total	109	104.79

Table 4.2.7a: Allele frequency at locus D16S539

Allele	Number of times appeared	Total number of times appeared, *x*	Allele frequency, *x*/218
8	2(1)+1+2+4+3	12	0.0550
9	1+2(3)+3+13+7+3	33	0.1514
10	2+3+2(4)+6+6+3+1	29	0.1330
11	4+13+6+2(12)+12+5+2	66	0.3028
12	3+7+6+12+2(6)+6+2	48	0.2202
13	3+3+5+6+2(3)+1	24	0.1101
14	1+2+2+1	6	0.0275
		$\sum = 218$	$\sum = 1.0000$

Table 4.2.7b: Expected number for each genotype at locus D16S539

Genotype	Observed number, O	Expected number, E
8 – 8	1	0.33
8 – 9	1	1.82
8 – 10	2	1.59
8 – 11	4	3.63
8 – 12	3	2.64
9 – 9	3	2.50
9 – 10	3	4.39
9 – 11	13	9.99
9 – 12	7	7.27
9 – 13	3	3.63
10 – 10	4	1.93
10 – 11	6	8.78
10 – 12	6	6.38
10 – 13	3	3.19
10 – 14	1	0.80
11 – 11	12	9.99
11 – 12	12	14.54
11 – 13	5	7.27
11 – 14	2	1.82
12 – 12	6	5.29
12 – 13	6	5.29
12 – 14	2	1.32
13 – 13	3	1.32
13 – 14	1	0.66
Total	109	106.37

Table 4.2.8a: Allele frequency at locus D7S820

Allele	Number of times appeared	Total number of times appeared, *x*	Allele frequency, *x*/218
6	1	1	0.0046
7	1+1+1	3	0.0138
8	1+1+2(3)+6+21+10+10	55	0.2523
9	6+2(2)+6+7+3+1	27	0.1239
10	21+6+2(2)+6+10+3	50	0.2294
11	1+10+7+6+2(3)+10	40	0.1835
12	1+10+3+10+10+2(1)+1	37	0.1697
13	1+3+1	5	0.0229
		$\sum = 218$	$\sum = 1.0000$

Table 4.2.8b: Expected number for each genotype at locus D7S820

Genotype	Observed number, O	Expected number, E
6 – 8	1	0.25
7 – 8	1	0.76
7 – 11	1	0.55
7 – 12	1	0.51
8 – 8	3	6.94
8 – 9	6	6.81
8 – 10	21	12.62
8 – 11	10	10.09
8 – 12	10	9.33
9 – 9	2	1.67
9 – 10	6	6.20
9 – 11	7	4.96
9 – 12	3	4.58
9 – 13	1	0.62
10 – 10	2	5.74
10 – 11	6	9.18
10 – 12	10	8.49
10 – 13	3	1.15
11 – 11	3	3.67
11 – 12	10	6.79
12 – 12	1	3.14
12 – 13	1	0.85
Total	109	104.90

Table 4.2.9a: Allele frequency at locus D13S317

Allele	Number of times appeared	Total number of times appeared, *x*	Allele frequency, *x*/218
8	2(3)+2+1+10+13+6+1	39	0.1789
9	2+2(2)+5+5+1	17	0.0780
10	1+6+2+1	10	0.0459
11	10+5+6+2(11)+14+3+1	61	0.2798
12	13+5+2+14+2(9)+7+2+2	63	0.2890
13	6+1+1+3+7+2(2)	22	0.1009
14	1+1+2	4	0.0183
15	2	2	0.0092
		$\sum = 218$	$\sum = 1.0000$

Table 4.2.9b: Expected number for each genotype at locus D13S317

Genotype	Observed number, O	Expected number, E
8 – 8	3	3.49
8 – 9	2	3.04
8 – 10	1	1.79
8 – 11	10	10.91
8 – 12	13	11.27
8 – 13	6	3.94
8 – 14	1	0.71
9 – 9	2	0.66
9 – 11	5	4.76
9 – 12	5	4.91
9 – 13	1	1.72
10 – 11	6	2.80
10 – 12	2	2.89
10 – 13	1	1.01
11 – 11	11	8.53
11 – 12	14	17.63
11 – 13	3	6.15
11 – 14	1	1.12
12 – 12	9	9.10
12 – 13	7	6.36
12 – 14	2	1.15
12 – 15	2	0.58
13 – 13	2	1.11
Total	109	105.63

The above raw data was processed by using computer software called DNA-View (Brenner, 2002) and the completed result was compiled in Table 4.3 as followed.

Table 4.3: Allele frequencies for 9 STR loci in Sikhs from Malaysia

Allele	CSF1PO (n = 109)	TPOX (n = 109)	THO1 (n = 109)	F13A01 (n = 109)	FESFPS (n = 109)	vWA (n = 109)	D16S539 (n = 109)	D7S820 (n = 109)	D13S317 (n = 109)
3.2	-	-	-	0.1330	-	-	-	-	-
4	-	-	-	0.0917	-	-	-	-	-
5	-	-	-	0.3486	-	-	-	-	-
6	-	-	0.2615	0.1972	-	-	-	0.0046	-
7	-	0.0138	0.1422	0.1697	0.0046	-	-	0.0138	-
8	-	0.4404	0.1330	0.0046	0.0092	-	0.0550	0.2523	0.1789
9	0.0138	0.1468	0.2339	-	0.0046	-	0.1514	0.1239	0.0780
9.3	-	-	0.1881	-	-	-	-	-	-
10	0.2064	0.1101	0.0413	-	0.2615	-	0.1330	0.2294	0.0459
11	0.2890	0.2523	-	-	0.3899	-	0.3028	0.1835	0.2798
12	0.4083	0.0367	-	-	0.2706	0.0046	0.2202	0.1697	0.2890
13	0.0734	-	-	-	0.0459	0.0092	0.1101	0.0229	0.1009
14	0.0046	-	-	0.0046	0.0138	0.1193	0.0275	-	0.0183
15	0.0046	-	-	0.0138	-	0.0963	-	-	0.0092
16	-	-	-	0.0367	-	0.2202	-	-	-
17	-	-	-	-	-	0.3073	-	-	-
18	-	-	-	-	-	0.1881	-	-	-
19	-	-	-	-	-	0.0459	-	-	-
20	-	-	-	-	-	0.0092	-	-	-
H	81.66	84.86	80.19	80.32	76.78	89.54	93.72	92.71	90.00
PE	0.4843	0.4937	0.6058	0.5860	0.4673	0.6027	0.5872	0.4862	0.5886
PD	0.8566	0.8465	0.9265	0.9199	0.8646	0.9271	0.9445	0.9112	0.9364
Chi	1.7182	8.8579	6.2850	8.4454	5.5832	7.0524	6.2314	11.0373	9.2823
($p < 0.05$)	(df 8)	(df 7)	(df 9)	(df 10)	(df 6)	(df 13)	(df 12)	(df 11)	(df 11)
CDP	0.9999999996 [CDP = 1 – Product of (1 – PD) = 1 – $(1 – PD_1)(1 – PD_2)(1 – PD_3)$ …… $(1 – PD_n)$]								

H: Heterozygosity; PE: Power of Exclusion; PD: Power of Discrimination; Chi: Chi Square; CDP: Cumulative Discrimination Power

5.0 DISCUSSION

From the results that I obtained, no significant deviation from Hardy-Weinberg expectation for all the 9 STR loci was observed at each locus based on the chi-square test value (Table 4.3).

There are all together 68 alleles observed for these 9 STR loci (Table 4.3), with frequencies ranging from as low as 0.0046 (CSF1PO alleles 14 and 15, F13A01 alleles 8 and 14, FESFPS allele 7, vWA allele 12, and D7S820 allele 6) to as high as 0.4404 (TPOX allele 8). They are further expressed in figures 5.1, 5.2 and 5.3 below:

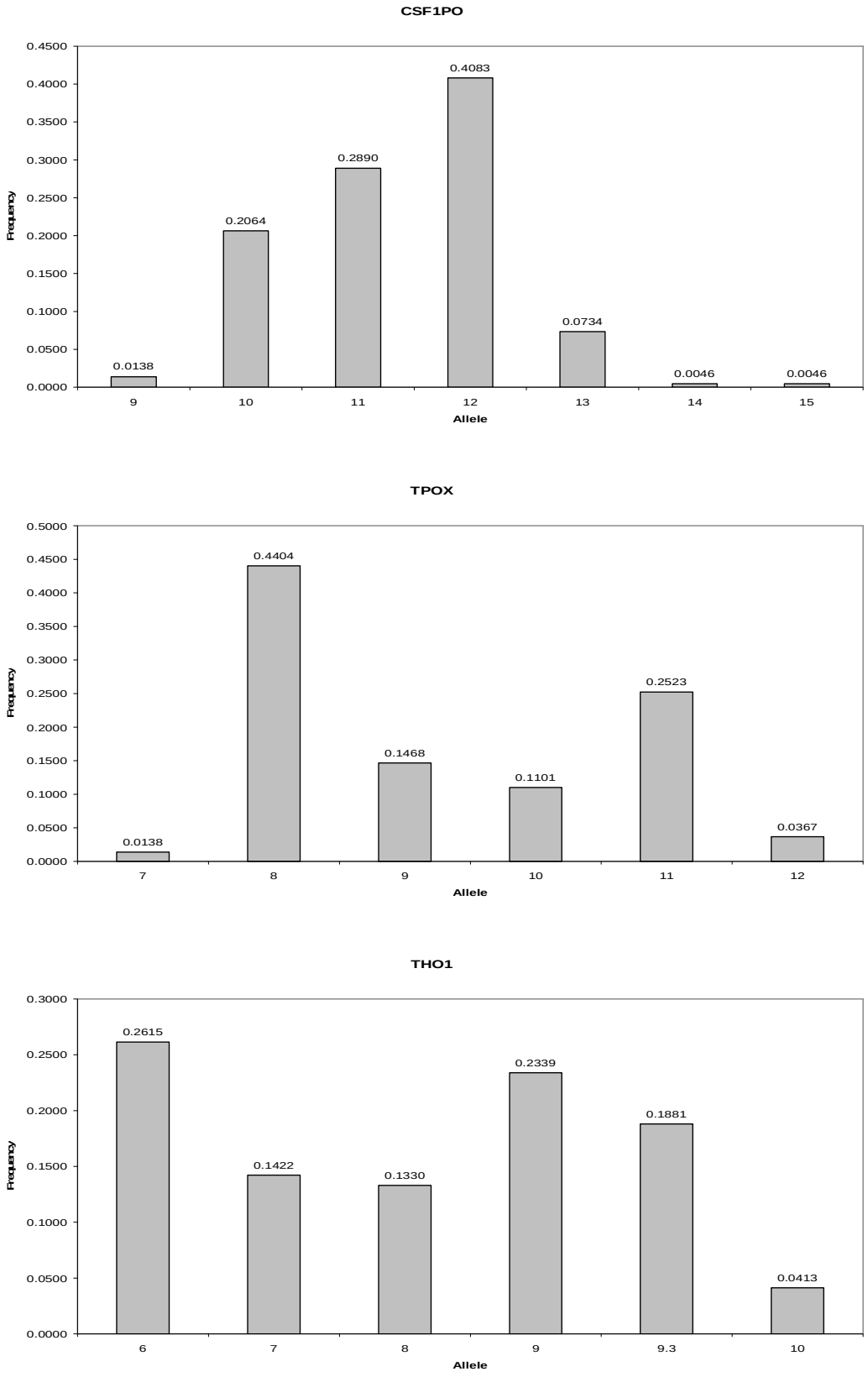

Figure 5.1: Allele frequencies for CSF1PO, TPOX and THO1.

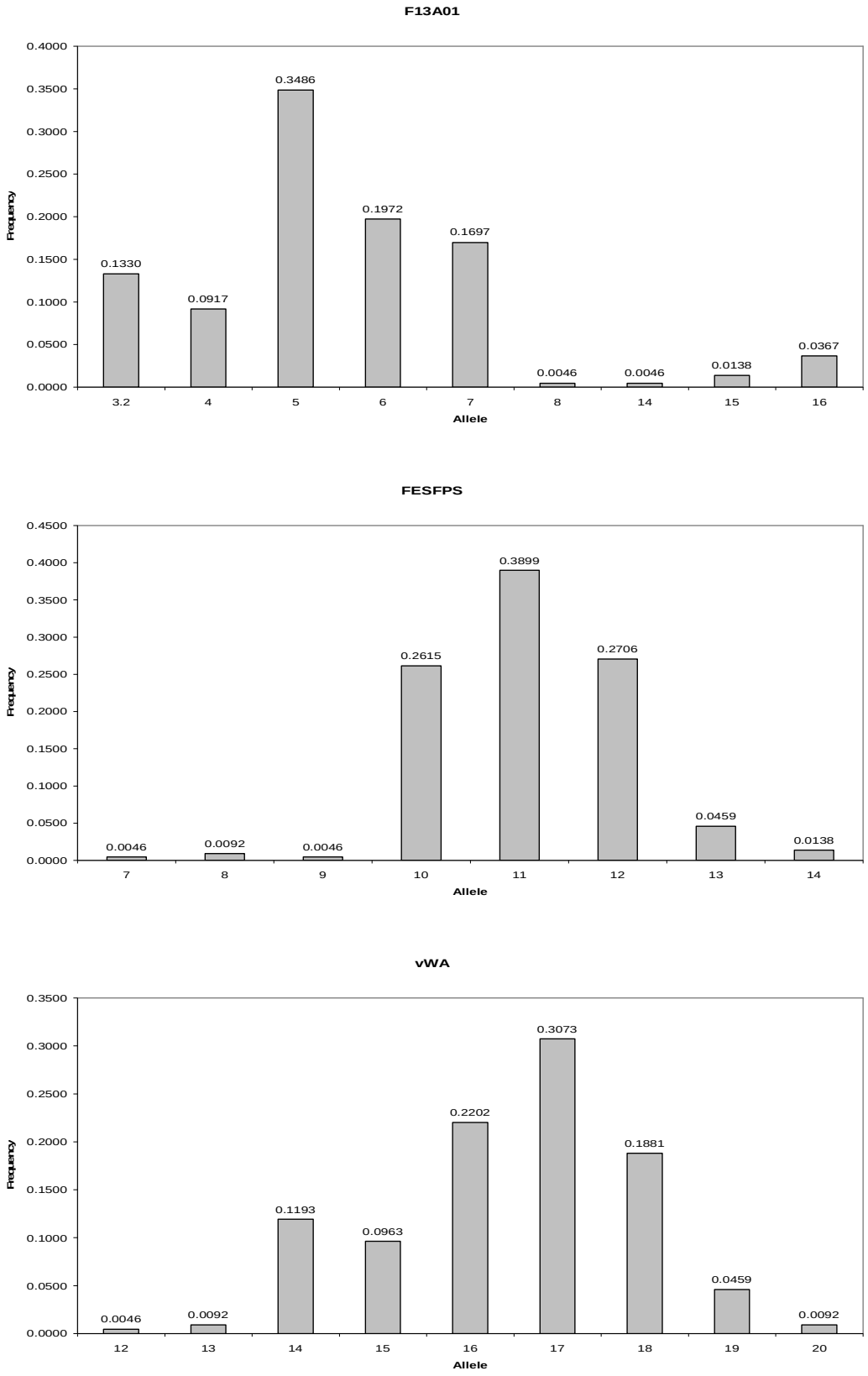

Figure 5.2: Allele frequencies for F13A01, FESFPS and vWA.

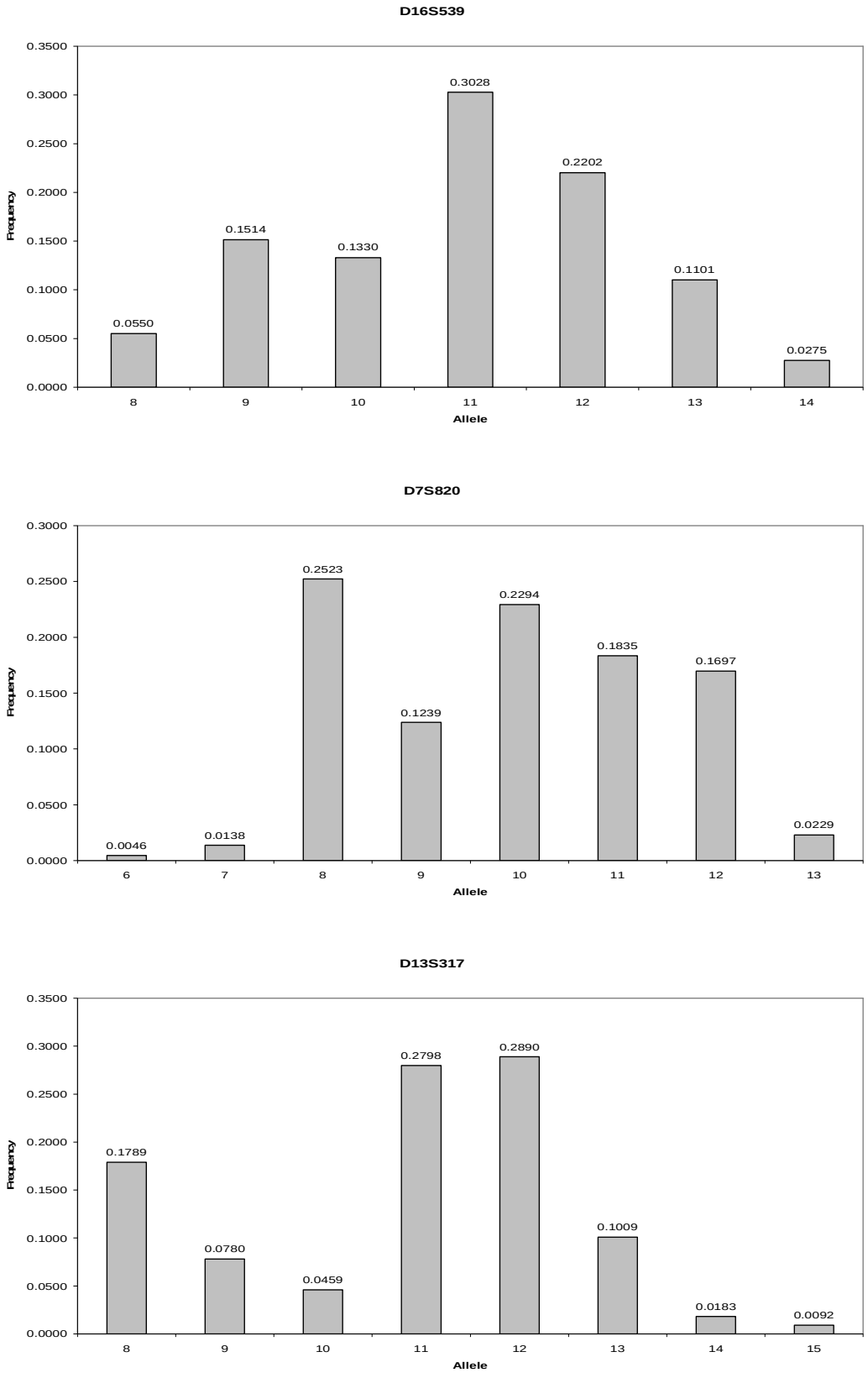

Figure 5.3: Allele frequencies for D16S539, D7S820 and D13S317.

Comparison of chi-square value obtained in Table 4.3 for each STR loci under respective degrees of freedom (df) suggests that the genotypes observed in 109 samples is closed to the expected number, based on the Hardy-Weinberg hypothesis. All the chi-square values calculated for each STR loci did not exceed the critical theoretical values, at $p = 0.05$ and 0.01, to which the observed chi-square is generally compared.

Next, from the result generated by the software DNA-View, the heterozygosity for my sample is high, with an average of 85.53, for the 9 STR loci. This again suggests that the DNA samples collected were segregated according to the Hardy-Weinberg hypothesis. This also suggests that the samples which I have collected were not from an inbred population. Thus, the individuals were sufficiently random and the samples were in Hardy-Weinberg equilibrium (Mpoloka et al., 2008).

Power of Exclusion (PE) is the probability that a randomly selected person would have an allele not detected in the specific profile, at any locus. In other words, PE is the ability of the test to exclude an individual. My data demonstrated a relatively high PE for each locus used, indicating the reliability of the markers as a genetic typing reference for the Malaysian Sikh community. The Cumulative Exclusion Power (CEP) for my data, combining all 9 STR loci used, gives a probability of 0.9992 which one can be confident that a suspect would not be falsely accused. The CEP is derived as shown:

Allele	Power of Exclusion (PE)	1- Power of Exclusion (1 – PE)
CSF1PO	0.4843	0.5157
TPOX	0.4937	0.5063
THO1	0.6058	0.3942
F13A01	0.5860	0.4140
FESFPS	0.4673	0.5327
vWA	0.6027	0.3973
D16S539	0.5872	0.4128
D7S820	0.4862	0.5138
D13S317	0.5886	0.4114
Product of (1 – Power of Exclusion)		7.87×10^{-4}

Cumulative Exclusion Power = 1 – [Product of (1 – Power of Exclusion)]

= $1 - 7.87 \times 10^{-4}$

= 0.999213

Power of Discrimination (PD) is the chance of picking someone at random who has a different allele numbers from another person in the same group at the examined loci. It is opposite to Matching Probability (Mpoloka et al., 2008). In other words, in criminal cases, PD is the ability of the test to pinpoint a suspect as the culprit of an offence. My data obtained a relatively high PD for each locus which confirms its reliability to be used as a reference database when a prosecution involves suspects from the Malaysian Sikh community. The Cumulative Discrimination Power (CDP) for my data, combining all 9 STR loci used, gives a probability of 0.9999999996 which one can be confident enough that when a match is obtained between a suspect and a crime scene sample, the suspect would be the culprit who committed the crime. The calculation of CDP is shown below:

Allele	Power of Discrimination (PD)	1- Power of Discrimination (1 – PD)
CSF1PO	0.8566	0.1434
TPOX	0.8465	0.1535
THO1	0.9265	0.0735
F13A01	0.9199	0.0801
FESFPS	0.8646	0.1354
vWA	0.9271	0.0729
D16S539	0.9445	0.0555
D7S820	0.9112	0.0888
D13S317	0.9364	0.0636
Product of (1 – Power of Discrimination)		4.00×10^{-10}

Cumulative Discrimination Power = 1 – [Product of (1 – Power of Discrimination)]

= $1 - 4.00 \times 10^{-10}$

= 0.9999999996

The CDP has to be expressed in such accuracy in order to eliminate the possibility that any two or more random individuals would have the same DNA profile that matches the crime scene sample. Of course, the more STR loci used, the more accurate the test would be. As for my data, chances that two randomly selected different Malaysian Sikh individuals having exactly the same DNA profile is 4.00×10^{-10} or 1 in 2,500,000,000. The estimated total Malaysian population, according to Department of Statistic Malaysia, updated on 21 June 2007, is 27.12million (*http://www.statistics.gov.my*). This 27.12 million Malaysians includes children and adults, regardless of gender, from various ethnic groups in Malaysia. The CDP of 1 in 2.5 billion means that if a suspect is a Malaysian Sikh and his DNA profile, tested using the 9 STR loci used in my project, matches a crime scene sample, based on the allele frequency database calculated in this project, it is almost impossible that there would be another person in Malaysia, Sikh or otherwise having exactly the same DNA profile as the suspect.

In order to further explain the use of the allele frequency database, I hereby present a simple example. A typical DNA case involves the comparison of two samples – an unknown or evidence sample, such as semen from a rape, and a known or reference sample, such as a blood sample from a suspect who is a Malaysian Sikh.

If the DNA profile obtained from the two samples are distinguishable (they "do not match"), that of course is obvious that the evidence sample does not come from the suspect. Hence, the suspect can be eliminated as a possible perpetrator. But, what if both the samples match with each other? If the DNA profile obtained from the two samples are indistinguishable (they "match"), that of course is evidence for the court that the samples have a common source – in this case, that the suspect contributed the semen.

How strong is the evidence? If the DNA profile consists of a combination of traits that are extremely rare, the evidence is very strong that the suspect is the contributor. To the extent that the DNA profile is not so rare, it is easier to imagine that the suspect might be unrelated to the crime and that he matches only by chance.

Therefore it is essential to have some idea as to the probability that a match would occur by chance. It is easiest to illustrate by example how the probability is determined:

DNA Profile		Allele frequency from Table 4.3		Genotype frequency for locus		Combined Matching Probability
Locus	Alleles	Database size of 218 alleles		Formula	Number	
CSF1PO	10	*p*	0.2064	2*pq*	0.1193	1 in 435
	11	*q*	0.2890			
TPOX	8	*p*	0.4404	p^2	0.1940	
	8					
THO1	6	*p*	0.2615	2*pq*	0.0984	
	9.3	*q*	0.1881			
F13A01	3.2	*p*	0.1330	p^2	0.0177	1 in 526316
	3.2					
FESFPS	8	*p*	0.0092	2*pq*	0.0008	
	13	*q*	0.0459			
vWA	16	*p*	0.2202	2*pq*	0.1353	
	17	*q*	0.3073			
D16S539	9	*p*	0.1514	p^2	0.0229	1 in 7142857
	9					
D7S820	12	*p*	0.1697	2*pq*	0.0078	
	13	*q*	0.0229			
D13S317	10	*p*	0.0459	2*pq*	0.0008	
	15	*q*	0.0092			
Profile Frequency					**6.2×10^{-16}**	**1 in 1.6×10^{15}**

Based on my Sikh database in Table 4.3, the allele 10 at the locus CSF1PO was observed 45 times in a Sikh population sample of 218 alleles (109 people). Therefore it is reasonable to estimate that there is a chance of p=0.2064 that any particular CSF1PO allele, selected at random, would be a 10. Similarly, the chance is about q=0.2890 for a

random CSF1PO allele to be 11 (allele 11 was observed 63 times in the same sample). Prior to typing the suspect, if we assume that he is not the donor of the evidence then we can think of him as someone who received a CSF1PO allele at random from each of his parents. The chance of receiving allele 10 from his mother and allele 11 from his father is therefore *pq*, and to receive 11 from mother and 10 from father is another *pq*, so the probability to be CSF1PO 10,11 by chance is 2*pq*. Hence about 12% (0.1193) of Sikhs in Malaysia have the 10,11 genotype at the CSF1PO locus.

At the TPOX locus, since both alleles are the same, there is only one term – *pp* or p^2, which represents the combined probability of inheriting the allele 8 from each parent. Hence about 19% (0.1940) of Sikhs in Malaysia have the same TPOX genotype as does the evidence. It is to be expected that the proportion of TPOX 8,8 people is still 19% even if attention is restricted only to people who have a particular CSF1PO genotype such as 10,11. Therefore the chance for a person to have the combined genotype in the two loci is 19% of 12% (0.1940 × 0.1193 = 0.0231), which is about 2%.

The calculations for the other 7 loci are similar, and taking them into account whittles the overall chance of a random person to have the combined genotype from 2% down to about $(6.2 \times 10^{-14})\%$, which is equivalent to a probability of $1/1.6 \times 10^{15}$. When using CTT Triplex alone for this case, the matching probability is 1 in 435; FFv Triplex alone is 1 in 526316; and SilverSTR III Triplex alone is 1 in 7142857. Combining all these Triplexes will give a matching probability as high as 6.2×10^{-16} in this case.

In summary, the probability of a particular multiple-locus genotype is obtained by multiplication – by multiplying together the frequencies of the per-locus genotypes,

which is to say, by multiplying together the frequencies of all the individual alleles and including in addition a factor of 2 for each heterozygous locus. This way to obtain the frequency of a DNA profile is called the Product Rule.

The profile frequency is sometimes referred to as the random match probability, or the chance of a random match. In the example case, the overall profile frequency is 6.2×10^{-16} or about $1/1.6 \times 10^{15}$. Therefore, a summary of the evidence is that:

> Either the suspect contributed the evidence, or an unlikely coincidence happened – the once-in-1.6×10^{15} (1.6 quadrillion) coincidence that an unrelated person would by chance have the same DNA profile as that obtained from the evidence.

It is obvious that such coincidence is very unlikely, or perhaps, almost impossible. This example demonstrates that DNA typing is truly an extremely accurate method in solving crimes or issues where DNA is involved.

Despite its accuracy, DNA typing is continuously being challenged in the court for its relevance. This is based on an argument that members from same ethnic group are more closely related than members from other ethnic group. In other words, in a population that contains groups with characteristic allele frequencies, knowledge of one allele in a person's genotype might carry some information about the group to which the person belongs, and in turn alters the statistical expectation for the other alleles in the genotype. For example, a person who has one allele that is common among Chinese is more likely to be of Chinese descent and is thus more likely to carry additional alleles that are common among Chinese. The true genotype frequency is thus higher than would be predicted by applying the multiplication rule using the average frequency in the entire population.

To illustrate this problem with a hypothetical example, suppose that a particular allele at a VNTR locus has a 1% frequency in the general population, but a 20% frequency in a specific subgroup. The frequency of homozygotes for the allele would be calculated to be 1 in 10,000 according to the allele frequency determined by sampling the general population, but would actually be 1 in 25 for the subgroup. This is a hypothetical and extreme example, but illustrates the potential effect of demography on gene frequency estimation.

The key question underlying the use of the product rule is whether the actual populations have significant substructure for the loci used for forensic typing. How can one address the possibility of population substructure? In principle, one might consider three approaches:

1) Carry out population studies on a large mixed population, such as a racial group, and use statistical test to detect the presence of substructure;
2) Derive theoretical principles that place bounds on the possible degree of population substructure;
3) Directly sample different groups and compare the observed allele frequencies.

The third offers the soundest foundation for assessing population substructure and is adopted in my study to compare the Sikh community to the three main ethnic groups (Malay, Chinese and Indian) in Malaysia. Tables 5.1 to Table 5.7 show the allele frequencies of different loci for the Malaysian Sikh, Malay, Chinese and India, whereas Figure 5.4 to Figure 5.10 show the distribution pattern of each allele at different loci based on allele frequencies in respective table.

Table 5.1: Allele frequencies for STR locus CSF1PO of other populations in Malaysia

Allele	Table 4.3	Lay Hong Seah et al., (2003)			Kong Boon Lim et al., (2000)		
	Sikh	Malay	Chinese	Indian	Malay	Chinese	Indian
6	-	-	-	-	-	-	-
7	-	-	0.0070	-	-	-	0.0038
8	-	0.0050	0.0020	-	0.0035	0.0067	0.0076
9	0.0138	0.0190	0.0460	0.0220	0.0106	0.0436	0.0265
10	0.2064	0.2050	0.2330	0.1600	0.2199	0.2651	0.1439
11	0.2890	0.3620	0.2740	0.3330	0.3191	0.2383	0.2917
12	0.4083	0.3330	0.3380	0.4190	0.3688	0.3356	0.4053
13	0.0734	0.0620	0.0940	0.0620	0.0532	0.1040	0.1023
14	0.0046	0.0070	0.0050	0.0050	0.0213	0.0034	0.0114
15	0.0046	0.0070	0.0020	-	0.0035	0.0034	0.0076
16	-	-	-	-	-	-	-
H	81.66	0.7670	0.6990	0.6790	0.6596	0.7651	0.7576
PE	0.4843	0.5390	0.4260	0.3970	0.4581	0.5156	0.4848
PD	0.8566	0.8450	0.8960	0.8430	0.8635	0.8935	0.8763

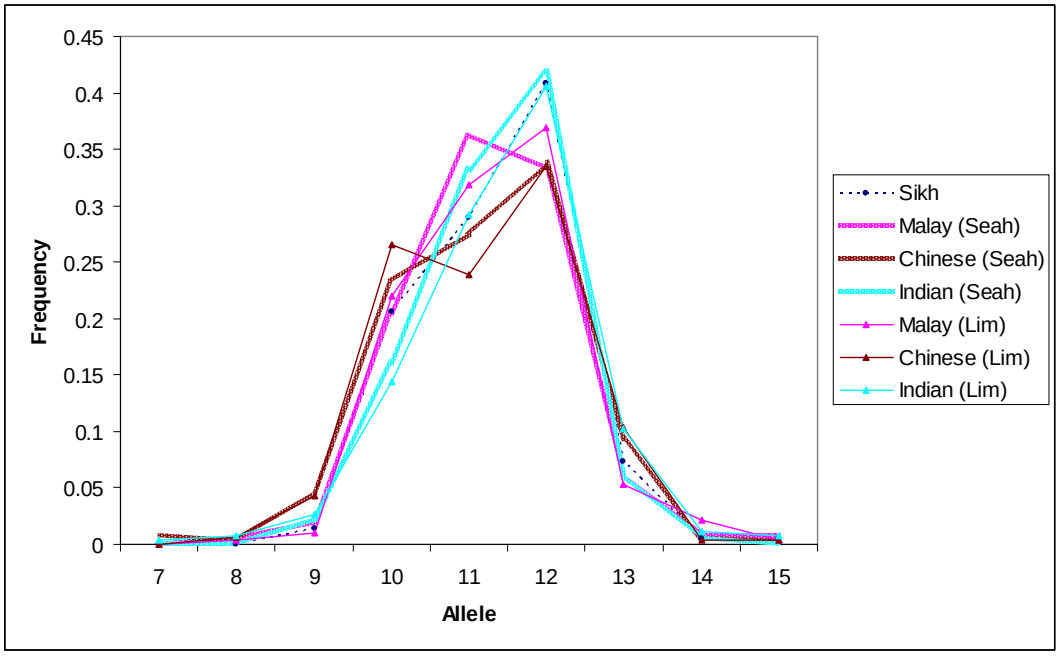

Figure 5.4: Distribution pattern of allele frequencies based on Table 5.1

Table 5.2: Allele frequencies for STR locus TPOX of other populations in Malaysia

Allele	Table 4.3	Lay Hong Seah et al., (2003)			Kong Boon Lim et al., (2000)		
	Sikh	Malay	Chinese	Indian	Malay	Chinese	Indian
6	-	-	-	-	-	-	-
7	0.0138	0.0020	-	-	0.0032	-	-
8	0.4404	0.5620	0.5620	0.3280	0.5353	0.5461	0.3358
9	0.1468	0.1170	0.1210	0.1720	0.1442	0.1414	0.1455
10	0.1101	0.0240	0.0180	0.1120	0.0481	0.0329	0.0709
11	0.2523	0.2670	0.2900	0.3560	0.2628	0.2566	0.4179
12	0.0367	0.0290	0.0090	0.0026	0.0064	0.0230	0.0299
13	-	-	-	0.0050	-	-	-
14	-	-	-	-	-	-	-
H	84.86	0.6190	0.5570	0.6990	0.5897	0.6250	0.6940
PE	0.4937	0.3140	0.2430	0.4260	0.3675	0.3633	0.4314
PD	0.8465	0.7800	0.7660	0.8770	0.8004	0.7962	0.8459

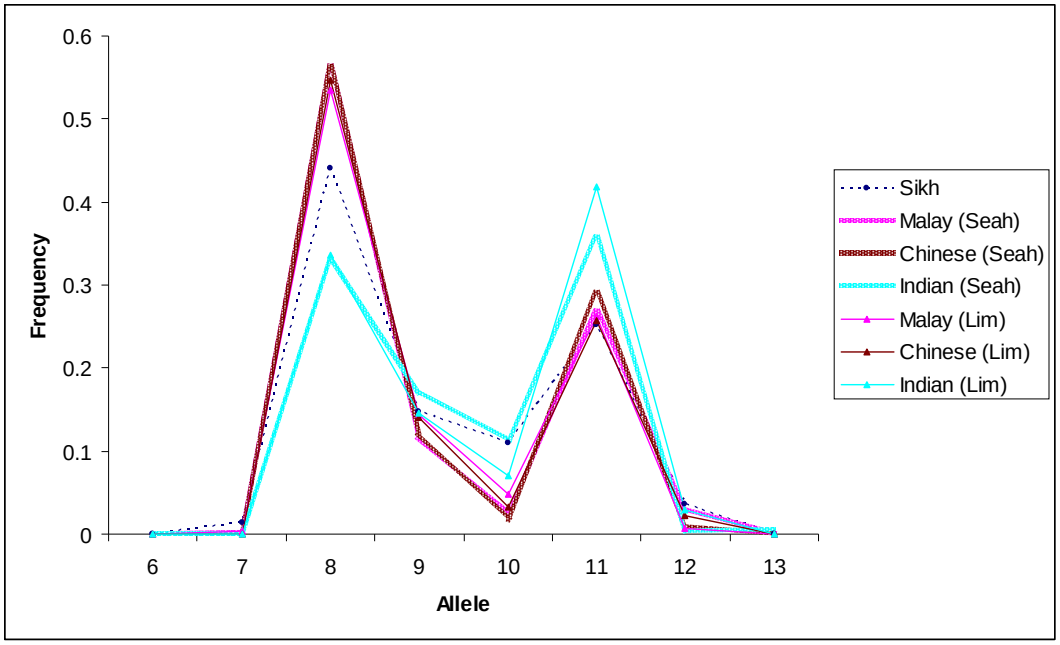

Figure 5.5: Distribution pattern of allele frequencies based on Table 5.2

Table 5.3: Allele frequencies for STR locus THO1 of other populations in Malaysia

Allele	Table 4.3	Lay Hong Seah et al., (2003)			Kong Boon Lim et al., (2000)		
	Sikh	Malay	Chinese	Indian	Malay	Chinese	Indian
4	-	-	-	-	-	-	-
5	-	0.0070	0.0050	-	-	-	-
6	**0.2615**	0.1330	0.1170	0.2700	**0.1323**	0.0105	0.2406
7	**0.1422**	**0.3260**	0.2980	0.1200	0.2323	0.2697	0.1541
8	0.1330	0.0830	0.0640	0.1340	0.1226	0.0658	0.1466
9	0.2339	0.3100	0.4590	0.3370	0.3452	0.4441	0.3158
9.3	0.1881	0.0830	0.0180	0.1240	0.0742	0.0395	0.1278
10	0.0413	0.0550	0.0370	0.0120	0.0935	0.0757	0.0150
11	-	0.0020	0.0020	0.0020	-	-	-
12	-	-	-	-	-	-	-
H	80.19	0.7430	0.7060	0.7990	0.8065	0.7105	0.8271
PE	0.6058	0.4980	0.4380	0.5970	0.5801	0.4809	0.5717
PD	0.9265	0.9040	0.8440	0.9040	0.9211	0.8731	0.9183

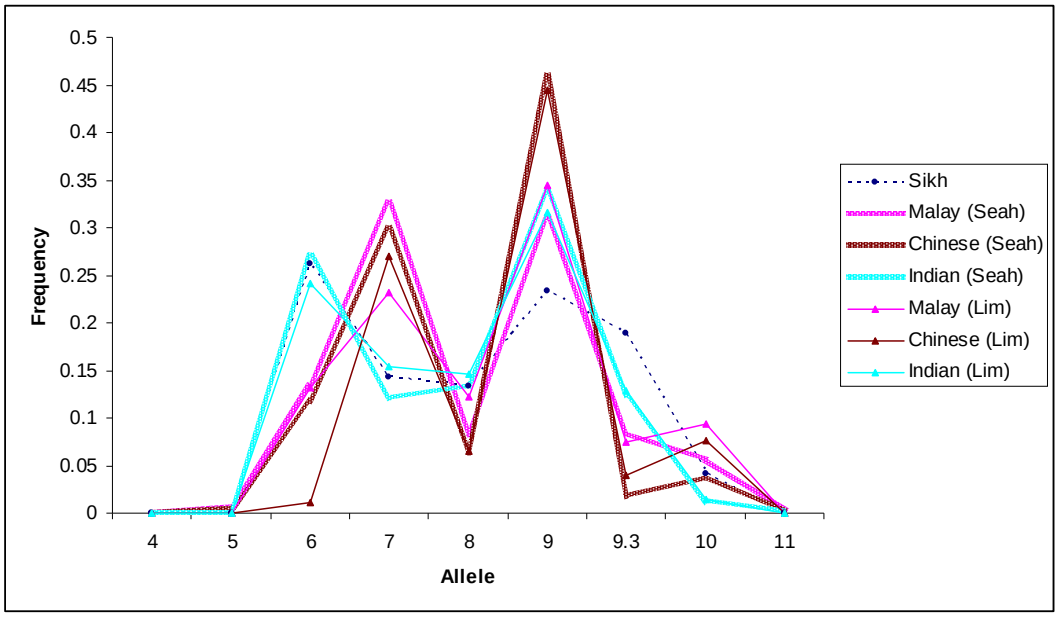

Figure 5.6: Distribution pattern of allele frequencies based on Table 5.3

Table 5.4: Allele frequencies for STR locus vWA of other populations in Malaysia

Allele	Table 4.3	Lay Hong Seah et al., (2003)			Kong Boon Lim et al., (2000)		
	Sikh	Malay	Chinese	Indian	Malay	Chinese	Indian
11	-	-	-	-	-	-	-
12	0.0046	-	-	-	-	-	-
13	0.0092	0.0020	-	0.0020	0.0032	0.0033	0.0111
14	0.1193	0.1860	0.2570	0.1440	0.2596	0.2451	0.1593
15	0.0963	0.0520	0.0390	0.1220	0.0609	0.0327	0.1000
16	0.2202	0.1810	0.1420	0.1990	0.1346	0.1242	0.2593
17	0.3073	0.2450	0.2480	0.2490	0.2372	0.2712	0.2148
18	0.1881	0.2290	0.2110	0.1580	0.2244	0.2288	0.1667
19	0.0459	0.0900	0.0800	0.1170	0.0705	0.0752	0.0704
20	0.0092	0.0140	0.0210	0.0100	0.0096	0.0196	0.0148
21	-	-	0.0020	-	-	-	0.0037
H	89.54	0.7480	0.7660	0.8090	0.8333	0.7908	0.8296
PE	0.6027	0.5060	0.5380	0.6150	0.6034	0.5908	0.6388
PD	0.9271	0.9330	0.9320	0.9440	0.9299	0.9251	0.9421

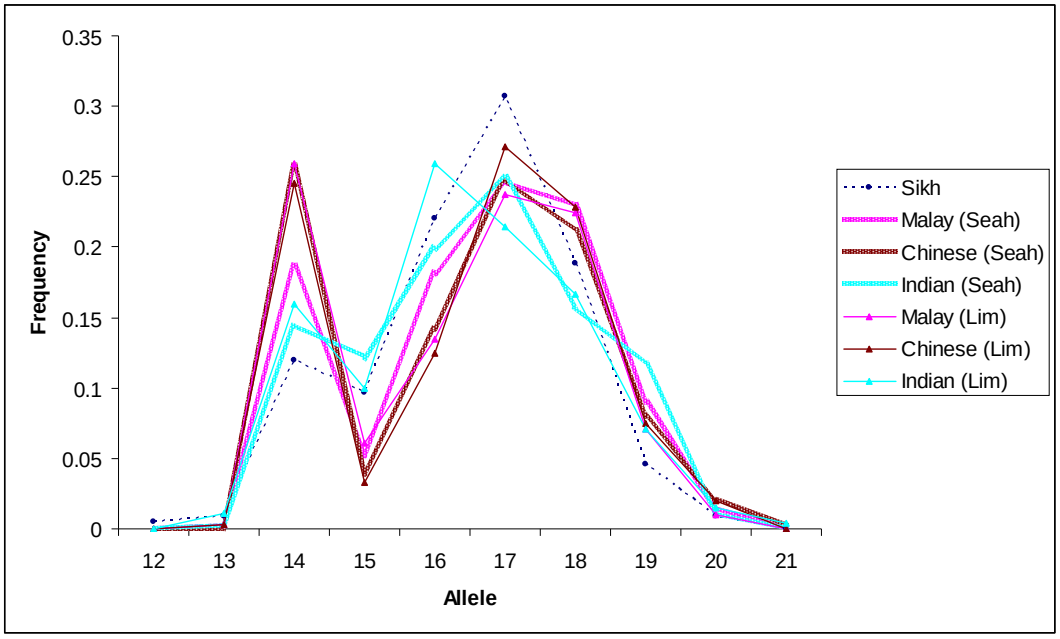

Figure 5.7: Distribution pattern of allele frequencies based on Table 5.4

Table 5.5: Allele frequencies for STR locus D16S539 of other populations in Malaysia

Allele	Table 4.3 Sikh	Lay Hong Seah et al., (2003)		
		Malay	Chinese	Indian
7	-	-	-	-
8	0.0550	0.0140	0.0050	0.0860
9	0.1514	0.1620	0.2350	0.1510
10	0.1330	0.1690	0.1300	0.0980
11	0.3028	0.3020	0.2720	0.3280
12	0.2202	0.2210	0.2170	0.2030
13	0.1101	0.1120	0.1320	0.1100
14	0.0275	0.0190	0.0070	0.0220
15	-	-	0.0020	0.0020
16	-	-	-	-
H	93.72	0.7520	0.7810	0.8180
PE	0.5872	0.5140	0.5640	0.6330
PD	0.9445	0.9240	0.9220	0.9290

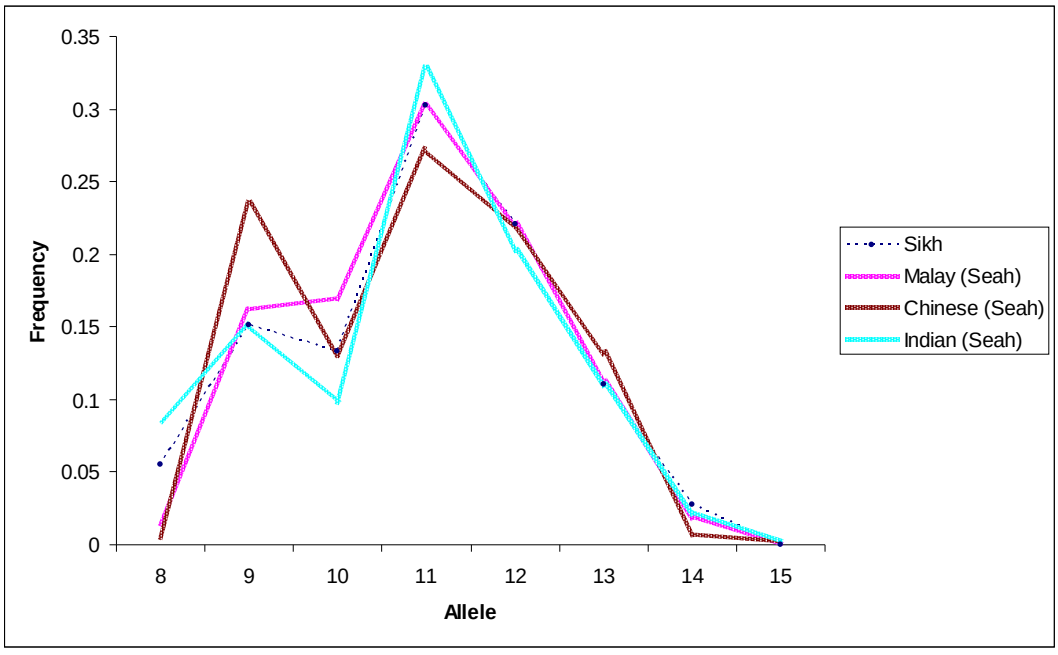

Figure 5.8: Distribution pattern of allele frequencies based on Table 5.5

Table 5.6: Allele frequencies for STR locus D7S820 of other populations in Malaysia

Allele	Table 4.3	Lay Hong Seah et al., (2003)			Kong Boon Lim et al., (2000)		
	Sikh	Malay	Chinese	Indian	Malay	Chinese	Indian
4	-	-	-	-	-	-	-
5	-	0.0020	0.0020	0.0020	-	-	-
6	0.0046	-	0.0020	-	-	-	-
7	0.0138	0.0120	0.0020	0.0500	0.0144	0.0033	0.0414
8	0.2523	0.2400	0.1070	0.2300	0.1942	0.1700	0.2444
9	0.1239	0.0550	0.0750	0.0550	0.0683	0.0533	0.0752
10	0.2294	0.2100	0.1850	0.2420	0.1583	0.1300	0.2368
11	0.1835	0.2880	0.3540	0.2250	0.3489	0.3533	0.2105
12	0.1697	0.1620	0.2370	0.1790	0.1871	0.2533	0.1729
13	0.0229	0.0260	0.0270	0.0170	0.0216	0.0367	0.0188
14	-	0.0050	0.0070	-	0.0072	-	-
15	-	-	-	-	-	-	-
H	92.71	0.7520	0.7810	0.8280	0.7410	0.7400	0.7744
PE	0.4862	0.5140	0.5640	0.6520	0.5695	0.5454	0.6073
PD	0.9112	0.9210	0.9090	0.9220	0.9169	0.9066	0.9314

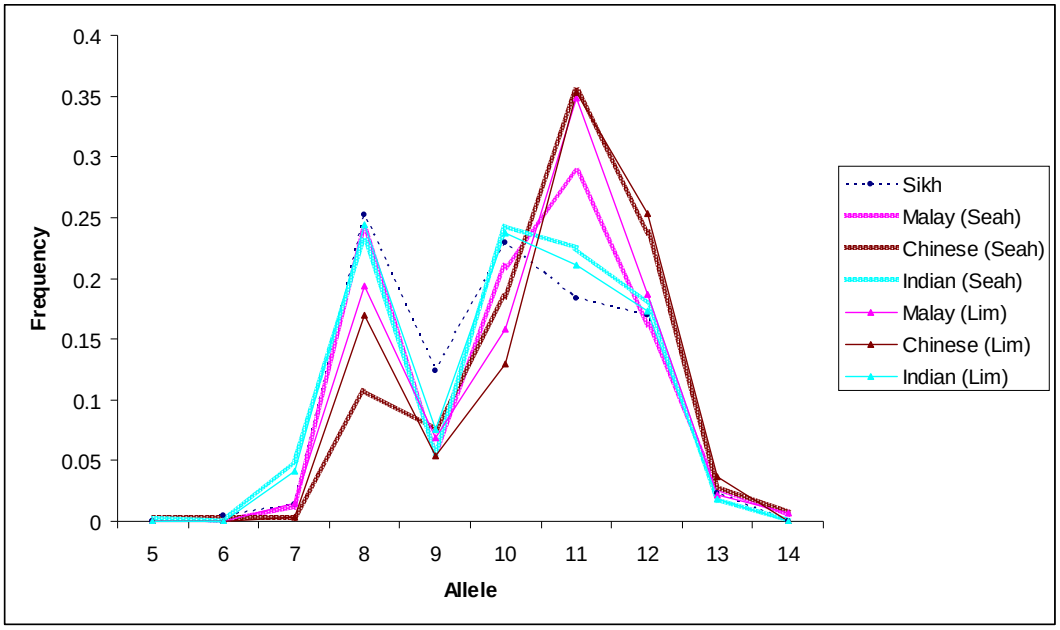

Figure 5.9: Distribution pattern of allele frequencies based on Table 5.6

Table 5.7: Allele frequencies for STR locus D13S317 of other populations in Malaysia

Allele	Table 4.3	Lay Hong Seah et al., (2003)			Kong Boon Lim et al., (2000)		
	Sikh	Malay	Chinese	Indian	Malay	Chinese	Indian
4	-	-	-	-	-	-	-
5	-	-	-	-	-	0.0033	-
6	-	-	-	-	-	-	-
7	-	0.0050	-	0.0140	-	-	0.0075
8	0.1789	0.2480	0.2790	0.2180	0.2411	0.2550	0.2015
9	0.0780	0.1400	0.1300	0.1240	0.1383	0.1358	0.1007
10	0.0459	0.1480	0.1460	0.0840	0.1277	0.1490	0.0448
11	0.2798	0.2400	0.2470	0.2320	0.2766	0.2815	0.2500
12	0.2890	0.1710	0.1550	0.2220	0.1773	0.1325	0.2948
13	0.1009	0.0400	0.0320	0.0740	0.0355	0.0397	0.0896
14	0.0183	0.0020	0.0110	0.0330	0.0035	0.0033	0.0112
15	0.0092	0.0050	-	-	-	-	-
16	-	-	-	-	-	-	-
H	90.00	0.7810	0.8170	0.7800	0.8440	0.7881	0.7687
PE	0.5886	0.5640	0.6320	0.5620	0.5990	0.5989	0.5903
PD	0.9364	0.9300	0.9230	0.9440	0.9286	0.9284	0.9248

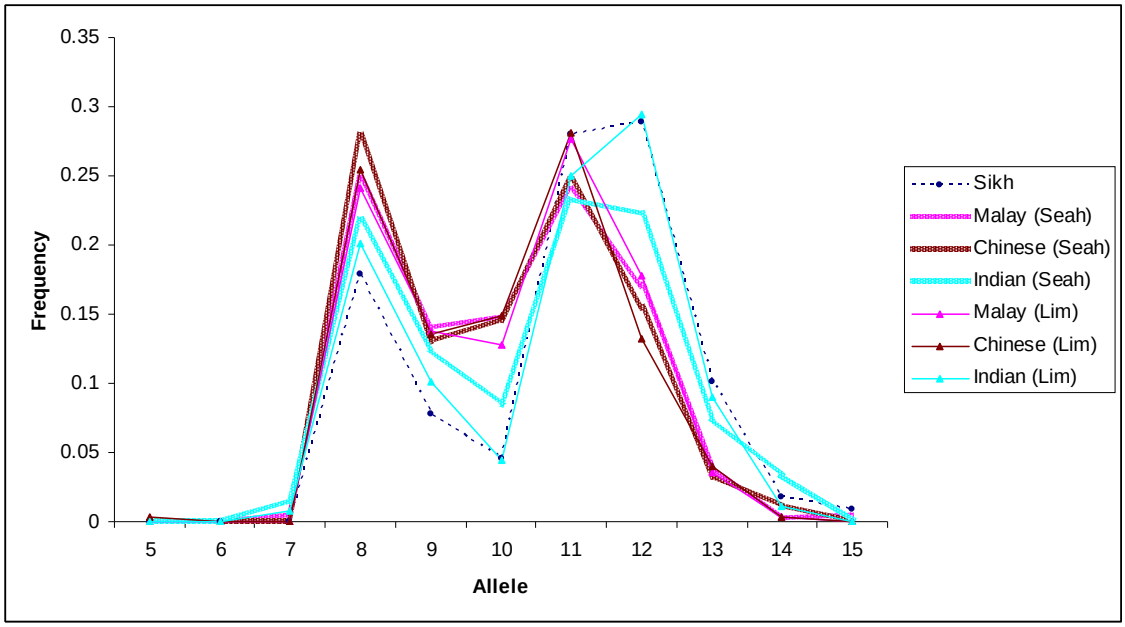

Figure 5.10: Distribution pattern of allele frequencies based on Table 5.7

As portrayed in the seven figures above (Figure 5.4 to Figure 5.10), not much difference in allele frequency among the Malay, Chinese, Indian and Sikh has been observed for all the loci analyzed. Similar patterns of distribution were observed for all the loci. This indicates that allele frequencies of the nine STR loci do not differ very much among the ethnic groups; hence, data for allele frequencies from the general population can also be used as reference when performing DNA typing analysis.

In order to further analyze genetic variation among ethnic groups, I performed a statistical calculation based on Hardy-Weinberg theory to compare the frequencies. I used data from Table 5.3 (numbers in bold) comparing allele 6 and 7 at locus THO1 between the Sikh and the Malay. The reason I have chosen these two alleles at this particular locus is because the frequency at these two alleles differs the most between the Sikh and Malay among the 7 loci which I compared above (see Table 5.1 – 5.7 and Figure 5.4 – 5.10). The allele frequency for THO1 6 is 0.2615 for the Sikh and 0.1323 for the Malay (Kong Boon Lim et al., 2000); THO1 7 is 0.1422 for the Sikh and 0.3260 for the Malay (Lay Hong Seah et al., 2003). Besides, I chose to compare the Sikh to the Malay is because of the huge difference in population size between these two ethnic groups in Malaysia. The Sikh is only about 3% whereas the Malay is about 51% of the total Malaysian population, which is roughly about 27 million (*http://www.statistics.gov.my*).

If genetic variation is significant among ethnic groups, comparison between the Sikh and the Malay should produce statistical significant result in the following examples.

Locus THO1	Allele 6	Allele 7
Sikh	0.2615	0.1422
Malay	0.1323 (Kong Boon Lim et al., 2000)	0.3260 (Lay Hong Seah et al., 2003)

	Proportion in overall Malaysian population (27million)	Number of population	Proportion in total population of Sikh and Malay (14.58million)
Sikh	3%	0.81million	6%
Malay	51%	13.77million	94%
Total	54%	14.58million	100%

If there is random mating within each group, Hardy-Weinberg equilibrium within the groups will produce these genotype frequencies:

Locus THO1	6,6	6,7	7,7
Sikh	0.0684	0.0744	0.0202
Malay	0.0175	0.0863	0.1063

In the overall Malaysian population, the observed genotype frequencies will be:

THO1	6,6	[(0.06)(0.0684) + (0.94)(0.0175)]0.54*	=	**0.0111**
THO1	6,7	[(0.06)(0.0744) + (0.94)(0.0863)]0.54*	=	**0.0462**
THO1	7,7	[(0.06)(0.0202) + (0.94)(0.1063)]0.54*	=	**0.0546**

* Multiplication of 0.54 is due to total population of the Sikh and Malay in overall Malaysian population is 54%.

Assuming there is no substructure among the Sikhs and Malays, the average allele frequencies, combining these two groups, will be:

THO1	allele 6	(0.06)(0.2615) + (0.94)(0.1323)	=	0.1401
THO1	allele 7	(0.06)(0.1422) + (0.94)(0.3260)	=	0.3150

This would correspond to the Hardy-Weinberg proportions of:

THO1	6,6	[(0.1401)(0.1401)]0.54*	=	**0.0106**
THO1	6,7	[2(0.1401)(0.3150)]0.54*	=	**0.0477**
THO1	7,7	[(0.3150)(0.3150)]0.54*	=	**0.0536**

* Multiplication of 0.54 is due to total population of the Sikh and Malay in overall Malaysian population is 54%.

As shown in the calculation above, even though there is substantial population substructure, the proportions do not differ greatly from Hardy-Weinberg expectation. This again showed that by analyzing specifically selected forensic application STR markers, such as the 9 STR loci used in my study, there is not much difference in allele frequencies across various ethnic groups. Therefore, DNA typing does not produce results that would bias the identity of a suspect to a certain ethnic group. However, more analysis on this matter should be done as the data pool grows over time. For the time being, it is still preferable that one uses data from specific ethnic group when the ethnicity of the individual in question is known.

Finally, I compared Heterozygosity (H), Power of Exclusion (PE) and Power of Discrimination (PD) of my data to the Malay, Chinese and Indian in Malaysia (Kong Boon Lim et al., 2000; Lay Hong Seah et al., 2003). Similar readings were observed across these 4 groups of ethnicity. This again showed that my data is in Hardy-Weinberg equilibrium and is relevant to be used as an allele frequency database.

6.0 CONCLUSION AND IMPLICATION

For the Sikh population in Malaysia which I have analyzed, no significant deviation from Hardy-Weinberg expectation for all the 9 STR loci was observed at each locus. Heterozygosity is high with an average of 85.53 for all the 9 STR loci. The Cumulative Discrimination Power (CDP) is 0.9999999996 or 99.99999996%. This also means that there is only a chance of 1 in 2,500,000,000 for two individuals having exactly the same DNA profile when analyzed with this 9 STR loci combination. The Cumulative Exclusion Power (CEP) is 0.999213 or 99.9213%, which means this system is able to exclude up to 999 people from a sample of 1000 people, leaving only one person whom DNA profile is a match. In both case, the CDP and CEP indicate that this system is reliable to pin point the correct person in the case of a match in STR typing test.

There is no huge difference in the pattern of allele distribution among the Sikh, Malay, Chinese and Indian in Malaysia. In other words, DNA typing method is unable to identify the distinct ethnic group of any sample examined but only identify the individual who left the traces at the genomic level, regardless of that particular individual's background. Therefore, this proved that DNA identification technology has no bias towards any external factor but targets solely on the human genome to discriminate a suspect from other innocents. Nonetheless, wherever possible, database from the same ethnic group of a suspect is recommended to be used.

7.0 APPENDIX

Multiplex System Information

GenePrint® STR Multiplex	Component Loci	Allelic Ladder Size Range (bases)	STR Ladder Alleles (number of repeats)	Other Known Alleles	K562 DNA Allele Sizes
CTT triplex	CSFIPO	295 – 327	7,8,9,10,11, 12,13,14,15	6	9,10
	TPOX	224 – 252	6,7,8,9,10, 11,12,13	None	8,9
	TH01	179 – 203	5,6,7,8,9, 10,11	9.3	9.3,9.3
FFv triplex	F13A01	283 – 331	4,5,6,7,8,9,10, 11,12,13,14,15,16	3.2,10	4,5
	FESFPS	222 – 250	7,8,9,10,11, 12,13,14	None	10,12
	vWA	139 – 167	13,14,15,16, 17,18,19,20	11,21	16,16
SilverSTR® III triplex	D16S539	264 – 304	5,8,9,10,11, 12,13,14,15	None	11,12
	D7S820	215 – 247	6,7,8,9,10, 11,12,13,14	None	9,11
	D13S317	165 - 197	7,8,9,10,11, 12,13,14,15	None	8,8

(Adopted from *GenePrint® STR System (Silver Stain Detection)*, Promega Technical Manual.)

Composition of Buffers and Solutions

(Adopted from *GenePrint® STR System (Silver Stain Detection)*, Promega Technical Manual.)

Saliva Swab DNA Extraction

1 M Tris-HCl pH 7.5, 1000 ml

121.1 g Tris base was dissolved in 800 ml of dH_2O and the pH was adjusted to pH 7.5 with saturated HCl. dH_2O was added to make it up to 1000 ml and then autoclaved.

0.5 M EDTA pH 8.0, 1000 ml

186.1 g $Na_2EDTA{\cdot}2H_2O$ was dissolved in 800 ml of dH_2O and the pH was adjusted to pH 8.0 with NaOH. dH_2O was added to make it up to 1000 ml and autoclaved.

5 M NaCl, 250 ml

73.05 g NaCl was dissolve in 200 ml of dH_2O and dH_2O was added to make it up to 250 ml.

20% SDS, 500 ml

100 g SDS was dissolved in 400 ml of dH_2O in an autoclave sterilized empty bottle and dH_2O was added to make it up to 500 ml.

Digestion buffer, 100 ml

1.0 ml of 1 M Tris-HCl (pH 7.5), 2.0 ml of 0.5 M EDTA, 10.0 ml of 20% SDS, 1.0 ml of 5 M NaCl and 86.0 ml of dH_2O were added together and autoclaved.

Proteinase K (20 mg/ml)

100 mg Proteinase K was dissolved in 5 ml dH_2O.

Chloroform-isoamyl (24:1), 250 ml

240 ml chloroform was added to 10 ml isoamyl alcohol in an Amber bottle.

3 M Sodium Acetate pH 5.2, 250 ml

102.025 g sodium acetate was dissolved in 200 ml dH_2O and the pH was adjusted to pH 5.2 with glacial acetic acid. dH_20 was added to make it up to 250 ml and autoclaved.

2 M Sodium Acetate pH 5.2, 300 ml

200 ml 3 M sodium acetate was added to 100 ml dH_20.

70% Ethanol, 500 ml

350 ml absolute ethanol was added to 150 ml dH_20.

TE buffer, 1000 ml

10 ml 1M Tris-HCl was mixed with 0.2 ml of 0.5 M EDTA. 989.8 ml dH_20 was then added and autoclaved.

Electrophoresis

10% NaOH, 1000 ml

100 g NaOH was dissolved in 1000 ml plain water.

0.5% Acetic Acid in 95% Ethanol, 50 ml

0.25 ml acetic acid was mixed with 49.75 ml absolute ethanol.

Bind Silane

3 µl bind silane (silver staining kit) was mixed with 2 ml 0.5% acetic acid in ethanol. It was prepared freshly in 1.5 ml microcentrifuge tube every time when needed.

20% Ammonium Persulfate, 1 ml

100 mg ammonium persulfate was dissolved in 0.5 ml dH_2O. It was prepared freshly in 1.5ml microcentrifuge tube every time when needed.

40% Acrylamide:Bisacrylamide (19:1), 1000 ml

380 g acrylamide and 20 g bisacrylamide was dissolved in 500 ml dH_2O and dH_2O was added to make it up to 1000 ml. It was prepared in Amber bottle and was kept in the fridge.

10× TBE Buffer, 1000 ml

107.8 g Tris base and 7.44 g EDTA ($Na_2EDTA{\cdot}2H_2O$) was dissolved in 800 ml dH_2O and 46 g boric acid was added. The pH was adjusted to pH 8 by adding the remaining of boric acid. dH_2O was added to make it up to 1000 ml and autoclaved.

0.5× TBE Buffer, 1000 ml

50 ml 10× TBE buffer was added to 950 ml dH_2O.

8.0 REFERENCES / BIBLIOGRAPHY

Brenner C. (2002) DNA View 2003: User Manual. *http://www.dnaview.com*. Last accessed: 18th July 2008.

Budowle B, Carmody G, Chakraborty R and Monson KL (2000) Source Attribution of a Forensic DNA Profile. Forensic Science Communications July 2000, Volume 2, Number 3, U.S.

Burke T, Dolf G, Jeffreys AJ, Wolff R (1991) *DNA Fingerprinting: Approaches and Applications.* Birkhäuser Verlag, Germany.

Chattopadhyay PK, Cleef S, Duelmer M, Henke L, Henke J (2000) *Human autosomal short tandem repeat types in Jat Sikhs from North India*. Forensic Science International 113: 29–32.

Department of Statistic Malaysia, *http://www.statistic.gov.my*. 18th July 2008.

Edwards A, Civitello A, Hammond HA, Caskey CT (1991) *DNA typing and genetic mapping with trimeric and tetrameric tandem repeats.* American Journal Human Genetics 49: 746–756.

Jeffreys AJ, Brookfield JFY, Semeonoff R (1985) *Positive identification of an immigration test-case using human DNA fingerprints.* Nature 317: 818–819.

Lim, K.B., Jeevan, N.H., Jaya, P., Othman, M.I., Lee, Y.H. (2001) *STR data for the AMPFℓSTR Profiler loci from the three main ethnic population groups (Malay, Chinese and Indian) in Malaysia.* Forensic Science International 119: 109–112.

Litt M, Luty JA (1989) *A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene.* American Journal Human Genetics 44: 397–401.

Mpoloka S.W., Kgotlele T. and Wally A. (2008) *Determination of allele frequencies in nine short tandem repeat loci of five human sub-populations in Botswana.* African Journal of Biotechnology Vol. 7 (8), pp. 1023–1027.

Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culver H, Martin C, Fujimoto E, Hoff M, Kumlin E, White R (1987) *Variable number of tandem repeat (VNTR) markers for human gene mapping.* Science 235: 1616–1622.

Payne JJ, Busuttil A, Smock W Ed (2003) *Forensic Medicine: Clinical and Pathological Aspects.* Bath Press Ltd, United Kingdom.

Pena SDJ, Chakraborty R, Epplen JT, Jeffreys AJ (1993) *DNA Fingerprinting: State of the Science.* Birkhäuser Verlag, Germany.

Promega, *Technical Manual: GenePrint® STR System (Silver Stain Detection),* Revised 7/06, USA.

Seah, L.H., Jeevan, N.H., Othman, M.I., Jaya, P., Ooi, Y.S., Wong, P.C., Kee, S.S. (2003) *STR Data for the AmpFℓSTR Identifiler loci in three ethnic groups (Malay, Chinese, and Indian) of the Malaysian population.* Forensic Science International 138: 134–137.

Tautz D, Renz M (1984) *Simple sequences are ubiquitous repetitive components of eukaryotic genomes.* Nucleic Acids Res 12: 4127–4138.

Vergnaud G, Mariat D, Zoroastro M, Lauthier V (1991) *Detection of single and multiple polymorphic loci by synthetic tandem repeats of short oligonucleotides.* Electrophoresis 12: 134–140.

Weber JL, May PE (1989) *Abundant class of human polymorphisms which can be typed using the polymerase chain reaction.* Am J Hum Genet 44: 388–396