UNIVERSITI SAINS MALAYSIA

Application of Polyacrylamide Gel Electrophoresis in Determining Y-Chromosomal Variations in the Malay Ethnic Group from Kelantan

Dissertation submitted in partial fulfillment for the Degree of Bachelor of Science (Health) in Biomedicine

Shazlin bt. Shaharudin

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

March 2006

CERTIFICATE

This is to certify that the dissertation entitled

"Application of polyacrylamide gel electrophoresis in determining Y-chromosomal variations in the Malay ethnic group from Kelantan"

is a bonafide record of research work done by

Miss Shazlin bt Shaharudin

Signature of Supervisor: .....

Name and address of Supervisor: Dr. Liza Sharmini bt Ahmad Tajudin
Department of Ophthalmology
School of Medical Sciences
Universiti Sains Malaysia
16150 Kubang Kerian, Kelantan

Signature of Supervisor: .....

Name and address of Supervisor: Dr. Zilfalil b Alwi
Director of Human Genome
School of Medical Sciences
Universiti Sains Malaysia
16150 Kubang Kerian, Kelantan

ACKNOWLEDGEMENT

I wish to extend my deepest appreciation and thanks first and foremost to my Supervisor, Dr Liza Sharmini Ahmad Tajudin and Dr Zilfalil Alwi, for their kind assistance and guidance ever since they took me under their wings. They relentlessly provided knowledge and basic skills I needed, regarding the thesis I'm working on and the science of molecular biology.

I would also like to thank the Science Officers, Technologists, Research Officers, Masters and Ph.D. students who were at the Human Genome Center, who helped a lot to solve the problems and doubts I had while working on this research project. My special thanks to Dr Hoh Boon Peng who sacrifice his time to help me. These individuals thought me a great deal on daily practices in molecular laboratory.

Finally, I would like to thank my parents, my best friends, and all the individuals mentioned above for giving me unconditional support and helping me to work through my final project. Thank you for keeping me going.

TABLE OF CONTENTS

LIST OF FIGURE

LIST OF TABLE

ABBREVIATION

APS	Ammonium Persulphate
bp	base pair
ddH_2O	Dionized Distilled Water
DNA	DeoxyRibonucleic Acid.
dNTPs	Deoxynucleotide Triphosphate
DYS	D = DNA: Y = Chromosome; S = Single copy sequence.
EDTA	Ethylenediaminetetraacetic acid
LB	Lithium Boric Acid
MtDNA	Mitochondrial DNA
MRCA	The Most Common Recent Ancestor
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
STR	Short Tandem Repeats
TBE	Tris-borate Acetate
TEMED	N,N,N,N-tetramethylethylenediamine

DEFINITION OF TERMS

DNA

DeoxyRibonucleic Acid. The genetic material of organisms, usually double-stranded; a class of nucleic acids identified by the presence of phosphate, deoxyribose (a sugar), and the four bases. Often forms the familiar double-helix. Within DNA are the code-words needed to form proteins, although much of the DNA is termed 'junk DNA' and has no known function.

DYS

D = DNA: Y = Chromosome; S = Single copy sequence. The DYS numbering scheme (e.g. DYS388, DYS390) for the Y-STR haplotype markers are controlled and administered by an international standards body called HUGO - Human Gene Nomenclature Committee - based at University College, London.

mtDNA

Mitochondrial DNA - The circular DNA contained inside the mitochondria. The mitochondria are small organelles residing in animal cells which provide the power to the cell. They occupy about one-fifth of each cell. The mtDNA is passed from mother to her offspring (both sons and daughters), but only the daughter will pass it on.The DNA sequence can be read and compared against a standard sequence (the Cambridge Reference Sequence) and deep (i.e. several thousands of years), but very broad genealogies can be deduced. A mtDNA haplogroup can usually be

assigned for any given sequence.

MRCA

The Most Recent Common Ancestor between two people. For example, for two 1st cousins, their shared grandparent is the MRCA. If the cousins were both boys, they would share their grandfather's Y-chromosome.

PCR

Polymerase chain reaction - A process carried out in a test tube that produces millions of copies of small sections of DNA. Using a heat resistant enzyme (DNA polymerase) and a mixture of other chemicals, cycles of hot and cold temperatures essentially photocopy a particular marker (or locus) of the DNA many times. Fluorescent tags are added to each copy so that they may be detected using laser analysers. A technique called multiplexing enhances the process.

STR marker

Short Tandem Repeats marker - A stretch of DNA where a small base sequence (usu. 2-6 base-pairs) repeats itself several times, giving a particular allele. For example, at the STR marker DYS391, the base sequence may read TCTA TCTA TCTA TCTA TCTA TCTA TCTA TCTA where TCTA is repeated eight times. Choosing markers that have been proven to have high variation between and within populations is desirable.

Y-chromosom

The sex chromosome that instructs a foetus to grow into a baby boy. It is passed down from generation to generation only through the male line i.e. from father to son, father to son, etc. It is approx. 60 million base-pairs long.

ABSTRACT

The application of polyacrylamide gel electrophoresis technique and microsatellite (or Y-chromosome Short Tandem Repeat, Y-STR) loci in population genetic study is reviewed. Malay population in Kelantan was surveyed in order to address hypotheses of their origin. We employed a set of three Y-STR markers to the non-recombining portion of the Y-chromosome to estimate the allele frequencies among Malay ethnic group from Kelantan population.

Five cc of blood samples were collected with informed consent of volunteers from the selected districts in Kelantan. We were able to identify nine paternally-unrelated males (for at least three to six generation) as our subjects. The DNA was extracted using Qiagen extraction kit, and three Y-STR markers that is DYS391, DYS392, and DYS393 were genotyped. The primers were sequenced and amplified. The amplified products were separated on 2% agarose gel and later with 8% non-denature polyacrylamide gel. The fragments were then visualized by UV light.

Identification of alleles were verified by comparing PCR product sizes with allelic ladders. Alleles frequencies were calculated. The Malay Kelantan have diversity both in terms of number of alleles and the evenness of their allele frequencies distribution. The observed distribution is clearly consistent with low levels of male gene flow between these patrilocal, Malay Kelantanese ethnic group.

Bigger sample size and more Y-STR loci examined could give complete analysis of Y-Chromosome substructure in this population.

ABSTRAK

Aplikasi teknik elektroforesis gel polyacrylamide dan mikrosatelit (atau "Y-chromosome Short Tandem Repeat", Y-STR) dalam pengajian genetik populasi telah diulas. Populasi Melayu di Kelantan telah dikaji untuk mencari hipotesis tentang asal usul mereka. Kami telah meletakkan satu set yang terdiri daripada tiga penanda Y-STR pada bahagian Y-kromosom yang tidak bereplikasi untuk menganggarkan frekuensi alel kumpulan etnik Melayu dari populasi Kelantan.

Lima cc sampel darah telah diperolehi dengan persetujuan daripada sukarelawan-sukarelawan dari seluruh daerah Kelantan. Kami telah berjaya mengenal pasti 9 lelaki tidak berhubungan darah antara satu sama lain (untuk sekurang-kurangnya tiga hingga enam generasi) sebagai subjek kami. Sampel darah tersebut telah diekstrak mengunakan kit Qiagen, dan tiga penanda Y-STR iaitu DYS391, DYS392, DYS393 telah digenotipkan. Primer-primer telah dirangkai dan diperbanyakkan ('sequenced and amplified"). Kemudian, produk tersebut telah diasingkan pada 2% gel agaros dan kemudiannya pada 8% gel polyacrylamide. Pecahan-pecahan itu, kemudiannya telah dilihat menggunakan cahaya UV.

Pengenalpastian alel telah dilakukan dengan membandingkan saiz PCR produk terhadap "ladder" alel. Alel frekuensi dikira. Variasi telah dikenalpasti dalam kalangan 9 sampel tersebut pada bahagian tertentu penanda "Y-STR". Melayu

kelantan didapati memiliki diversiti dari segi bilangan alel dan alel frekuensi. Walaupun sampel saiz kecil, namun penyebaran alel didapati konsisten dengan nilai aliran gen yang rendah dalam populasi Melayu Kelantan yang bersifat 'patrilocal'.

Usaha boleh dilakukan untuk mecipta semula sejarah generasi Melayu Kelantan melalui "paternal lineages", dengan meneliti perbezaan antara penanda Y-kromosom. Penggunaan sampel saiz yang lebih besar dan "Y-STR" yang lebih banyak juga adalah disarankan untuk mendapat profil struktur Y-kromosom bagi populasi ini secara keseluhurannya.

1 Introduction

1.1 Population Genetic

Population genetics is the study of the frequency of occurrence of alleles within and between populations. The information on the frequency can be applied to a variety of population issues such as understanding the genetic basis and probabilities for disease transmission, developing breeding programs for endangered or agricultural species, and elucidating the evolutionary history of a species (Jobling et at., 1995). With advances in the human genome project and the increasing availability of DNA markers scattered throughout the genome such as single nucleotide polymorphisms, variable number tandem repeats, and short sequence repeat polymorphisms, it has become increasingly possible to search for the genetic basis of complex human diseases using genomic wide screening methods (Jorde et al., 2001).

Traditionally, the study of population genetics involve the identification of different alleles through observation of the expressed traits, physical manifestation of a gene, called the phenotype. Mendelian genetics allow us to identify the heritable form of a gene (genotype) including individual variants (alleles). Advances in molecular genetics facilitated identification of single genes at the molecular or biochemical level (Underhill, 2003). Regardless of the method used to identify genes and their alleles, we use statistical analyses of allele frequencies to understand and make predictions about gene flow in populations.

1.2 **Deoxyribonucleic Acid (DNA)**

DNA is a long, double chain of subunits called nucleotides or bases. There are four different bases: adenine (A), guanine (G), cytosine (C) and thymine (T). The two chains line up next to each other to make a DNA molecule, adenine (A) and thymine (T) pairing each other as do guanine (G) and cytosine (C). DNA molecule is characterized by the sequence of these base pairs (bp) along the chain. It is this sequence that codes for all of an organism's traits (Jobling et al, 1995).

Mutations and the rearrangement of maternal and paternal genes through sexual reproduction, ensure that each member of a species has a unique DNA sequence. The ideal way to distinguish an individual from other person is to describe the entire sequence of nucleotides in his or her DNA. However, each human genome is made up of more than 3 billion nucleotide base pairs, describing an individual's complete DNA would be too complicated and not practical (Hammer, 1995).

1.3 **Y-Chromosome**

The human Y chromosome responsible in the development of the testis, which determine maleness of a person. This little chromosome (as shown in Figure 1.1), carries about 2% of a father's genetic contribution to his sons and programmes the early embryo to develop as a male.

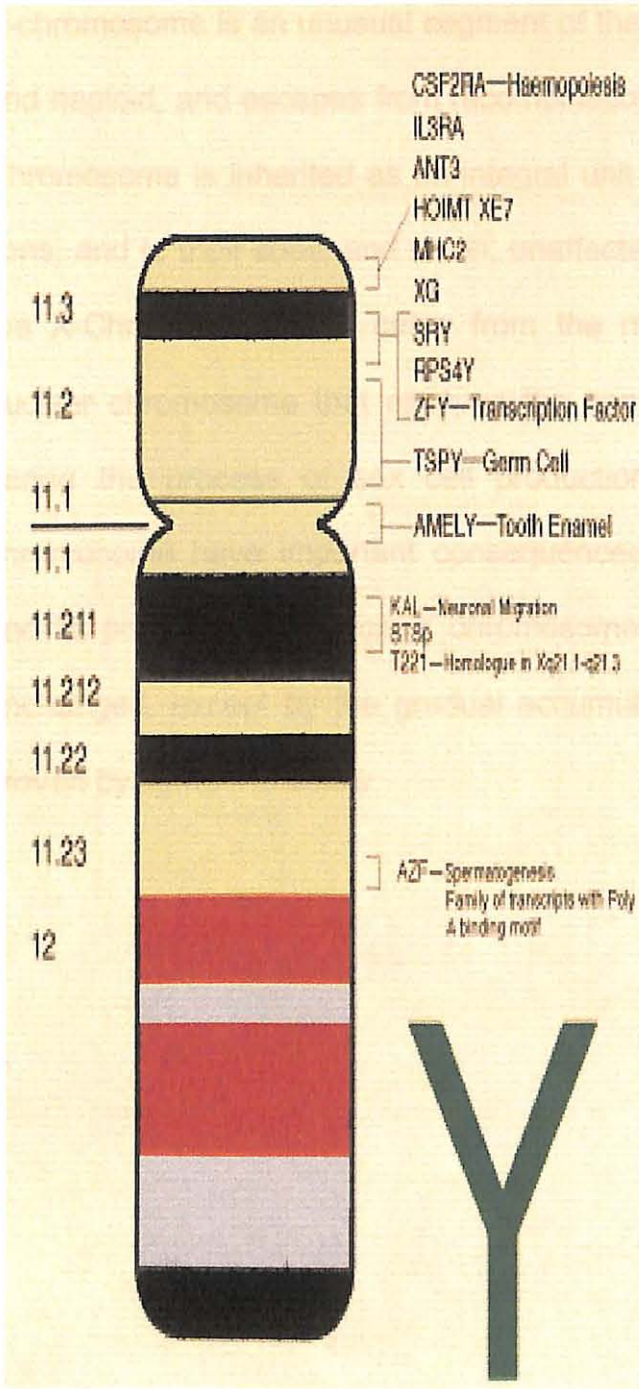

Figure 1.1 : Y-chromosome

Y-chromosome is an unusual segment of the human genome since it is male-specific and haploid, and escapes from recombination (Seielstad et al., 1999). Most of the Y-Chromosome is inherited as an integral unit passed without alteration from father to sons, and to their sons, and so on, unaffected by exchange or any other influence of the X-Chromosome that came from the mother (Underhill, 2003). It is the only nuclear chromosome that escapes the continual rearrangement of parental genes during the process of sex cell production. These unique properties of the Y-chromosome have important consequences for its mutation processes, its genes, and its population genetics: Y chromosomes pass down from father to son largely unchanged, except by the gradual accumulation of mutations. This statement was proven by figure 1.2 below.

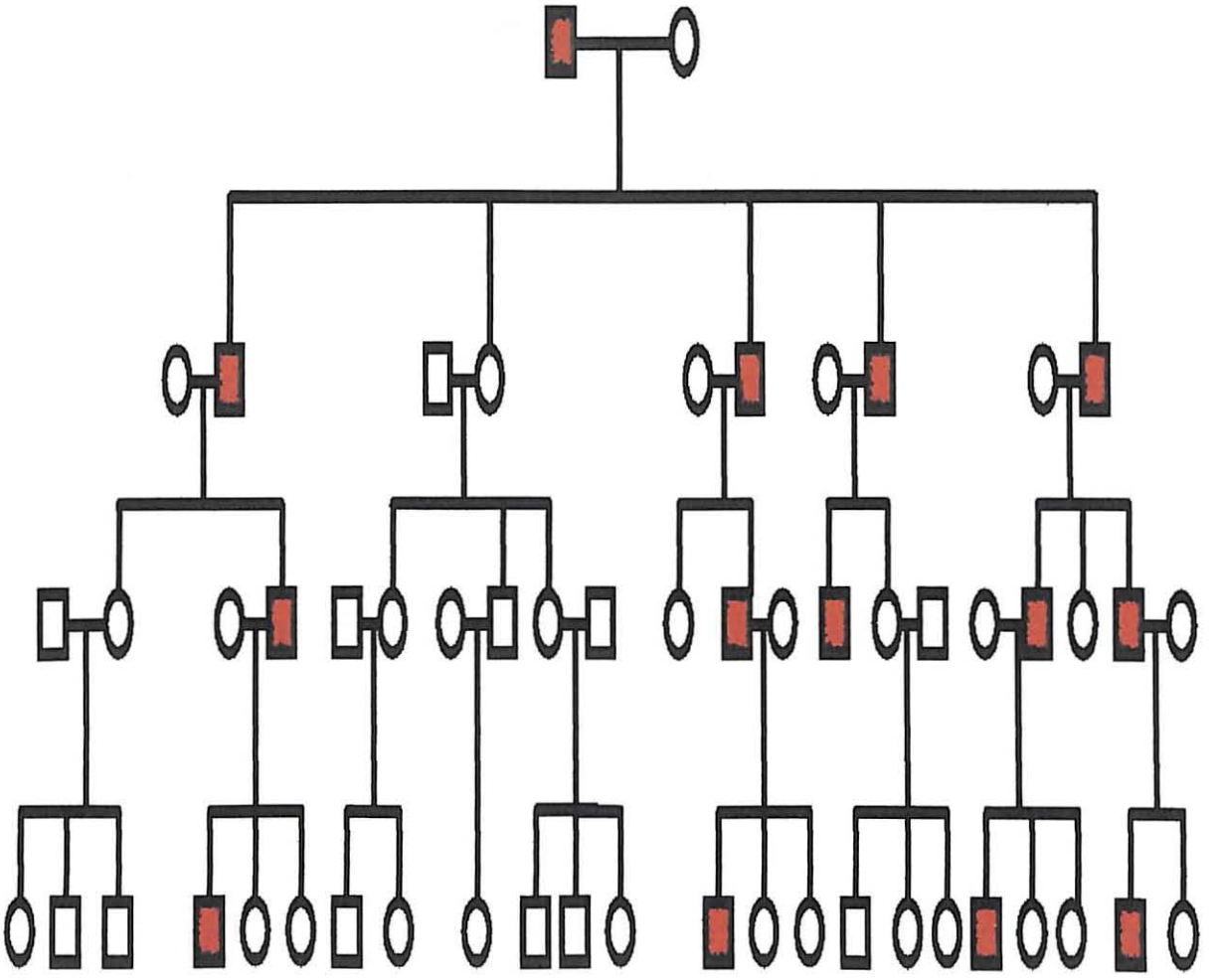

Figure 1. 2: Uninterrupted male-male link for four generations

- ■ Male sharing Y chromosome via an uninterrupted male-male link with common ancestor
- □ Male with Y chromosome from outside the paternal line
- ○ Females--have no Y chromosome

We can attempt to reconstruct a history of human paternal lineages by examining the differences between modern Y chromosomes - as DNA polymorphisms (Jobling et al.,1995). This complements maternal lineage studies using mitochondrial DNA (mtDNA) and studies using biparentally inherited markers in the rest of the genome. We can apply these to questions of population structure and history, genealogy, forensics, and the investigation of selective influences on the Y chromosome (Jorde et al., 2001).

1.4 **Simple Tandem Repeats**

Scientists had looked for sets of nucleotide sequences that are highly polymorphic, that is a section of DNA where a variety of different sequences are found among individuals in the same human population. These sets are referred to as markers and are and usually reflect some coding of scientific data from the laboratory. Only about five percent of human DNA is thought to code for traits. Most of the rest is made of long, apparently nonfunctional, sometimes referred to a "junk DNA". Within these nonfunctional stretches are short, moderately repetitive base pair sequences (Gonser et al., 2000). The number of repeats is inherited and is easily detectable making them ideal identifying markers. The number of repeating units can occasionally change during evolution and descent. Thus, they are useful markers for familial relationships and have been used in paternity testing, forensic science and in the identification of human remains.

There are two types of repetitive sequences, that is variable number tandem repeats (VNTR) and short tandem repeats (STR). Variable number tandem repeats (VNTR) are repeated sequences that typically range from 10 to 80 bps. These occur quite frequently in the human genome but there are relatively few different types. Short tandem repeat (STR) sequences, sometimes called microsatellites, are much shorter (2-6 bps) and may be repeated as many as 100 times in a head-tail manner at a given location on a chromosome (Edwards et al, 1991). The 16 base pair sequence of "gatagatagatagata" would represent 4 repeats of the sequence "gata". These repeats are referred to as allele. The variation of the number of repeats of each marker enables discrimination between individuals. The human genome contains hundreds of thousands of these STRs all evenly distributed on all the chromosomes. STRs represent ideal markers for genetic typing because of their rich diversity, wide distribution, and polymorphism. With the availability of PCR, STR became a popular DNA marker since they were easy to amplify and highly variable between individuals (Zafarina, 2004).

1.4.1 **Y-Chromosome Simple Tandem Repeat (STR) Marker**

The Y-Chromosome has definable segments of DNA with known genetic characteristics. These segments are known as markers. These markers occur at an identifiable physical location on a chromosome known as a locus. Each marker is designated by a number (known as DYS#), according to international conventions (Seielstad, 1999). Technically the marker is what is tested and the locus is where the marker is located on the chromosome.