GENETIC RELATIONSHIP AMONG THE THREE NEGRITO (SEMANG) TRIBES OF PENINSULAR MALAYSIA: MICROSATELLITE ANALYSIS

by

MOHD REDHUAN BIN MOHD NOOR

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

March 2008

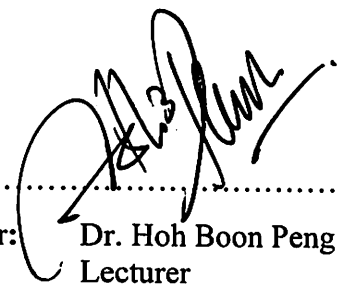
CERTIFICATE

This is to certify that the dissertation entitled

"Genetic Relationship among the Three Negrito (Semang) Tribes of Peninsular Malaysia: Microsatellite Analysis"

is a bonafide record of research work done by

Mr. Mohd Redhuan Bin Mohd Noor



Signature of supervisor:.....

Name and address of supervisor: Dr. Hoh Boon Peng
Lecturer
Human Genome Centre
School of Medical Sciences
Health Campus
16150 Kubang Kerian, Kelantan

ACKNOWLEDGEMENT

I would like to thank to God for his blessing in giving the strength and patience to accomplish this final year project. This research project was made possible with the cooperation and help of so many people.

I would like to thank my project supervisor, Dr. Hoh Boon Peng for his guidance and providing me with information during the study. Many thanks to the Director of Human Genome Centre, Assoc. Prof. Dr. Zilfalil Alwi for providing me the laboratory facilities throughout the project. Special thanks go to Ms. Nur Shafawati Rajab her endless help and assistance rendered. I would like to acknowledge staff nurse Mrs. Wey and Mrs. Zaliha for their help during sample collection.

Last but not least, I wish to thank to all my friends and family for their help, support and encouragement given. Thank you.

ABSTRACT

The struggle to determine susceptibility genes for complex disorders has stimulated geneticists to develop new approaches. The genetic markers rapidly superseded by microsatellites since they prolifically throughout the genome and are highly variable in repeat length and polymorphisms. The Negritos, believed to be the first aborigines arrived at Peninsular Malaysia, somewhat 10,000 years ago was thought to be related to Andaman islanders and Nicobar Islands. They are divided into six groups, namely Kensiu, Bateq, Mendriq, Jahai, Lanoh and Kintak. In this study, an attempt was made to optimize the locus D19S220 for GeneScan Analysis purpose. Marker was selected from Genethon Microsatellite Map (http://www.genlink.wustl.edu/genethon_frame/). DNA was extracted from the samples collected from three Negrito Tribes in Kelantan: Jahai, Mendriq and Bateq. A series of PCR optimization were done and run on the ABI 3100 capillary based genetic analyzer. PCR was then amplified on the Mendriq population in order to examine the PCR condition optimized.

ABSTRAK

Kepayahan untuk menentukan keupayaan sesuatu gen pada penyakit kompleks telah merangsang para ahli genetik untuk membangunkan satu pendekatan baru iaitu penanda genetik. Penanda genetik telah berkembang pesat terutama penanda genetik mikrosatelit kerana ia mempunyai nilai yang tinggi, panjang jujukan yang berulang dan tahap polimorfisma yang tinggi. Suku kaum Negrito dipercayai populasi pertama yang menduduki Semenanjung Malaysia 10,000 tahun yang lalu. Kaum ini terbahagi kepada enam kumpulan iaitu Kensiu, Jahai, Bateq, Lanoh, Mendriq dan Kintak. Kajian ini berusaha untuk mengoptimalkan tindak balas rantaian polimerase (PCR) bagi lokus D19S220 untuk analisis DNA/`GeneScan Analysis`. Lokus ini dipilih melalui laman web Genethon Microsatellite Map (http://www.genlink.wustl.edu/genethon_frame/). DNA diekstrakkan daripada darah individu dari tiga suku kaum terbanyak di Kelantan: Jahai, Mendriq dan Bateq. Proses optimisasi dan ujian PCR dilakukan dan analisis GeneScan dilakukan dengan *'ABI3100 capillary based genetic analyzer'*. Keadaan optimal PCR kemudian diuji pada populasi Mendriq untuk pengesahan.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1.0 INTRODUCTION

The struggle to determine susceptibility genes for complex disorders has stimulated geneticists to develop new approaches. During the last decade, a large number of approaches were proposed following the completion of human genome project. One of the approaches is to study on genetically isolated populations such as the aborigine populations or locally known as *Orang Asli* rather than the general populations. These homogenous populations have been proposed as a possible alternative because the exposure of environmental variation may be lower and thus the genetic make-up is expected to be less complex owing to founder effects.

Various markers have been identified and used in different aspect of human genetic studies. Classical genetic markers such as blood groups and protein level markers, i.e. isoenzymes had been available for decades. However, they have a relatively small degree of genetic divergence in existing human populations.

Later, DNA level genetic markers were use in attempt to sampling the variation at of the human genome, employing restriction fragment length polymorphism (RFLPs) and analysis of these data produced the pattern of population relationships similar to that obtained with classical markers. During the late 90's, genetic markers rapidly superseded by microsatellites since they are distributed throughout the genome and are highly variable in repeat length and polymorphisms. Hence, they had become useful tools for the genetic mapping of disease susceptibility loci (Ikari *et al.*, 2001).

1.1 Microsatellite Markers or Short Tandem Repeat (STR)

Microsatellites or Short Tandem Repeat (STR) are tandemly repeated DNA sequences containing repeat units range from two to six base pair in length eg. (CACACA...)n and can be amplified using polymerase chain reaction (Ruitberg *et al.*, 2000).

The microsatellite markers are of interest to many researchers studying in the population diversity as they are more polymorphic and hence more informative. They typically having more than ten alleles compared to other types of genetic markers. This is because the repetition involves few thousand base pairs and often many alleles present at the microsatellite locus are abundant, highly polymorphic and technically very simple to analyze (Payne, 1997).

Microsatellite markers have been used in search for susceptibility loci for many diseases. These markers offer a great statistical power for the data analysis either by parametric analysis of large families or non parametric analysis of sib-pairs. During recent decades, microsatellite has become one of the most popular sources of genetic markers. In human many thousands microsatellites markers have been identified and these marker are distributed over virtually every section of every chromosome (Chambers *et al.*, 2000). It has been used in forensics applications (Ruitberg *et al.*, 2000); paternity testing and in analysis of population structure (Payne, 1997).

Besides that, they have been an essential tool for genetic linkage analyses (Ikari *et al.*, 2001). Microsatellites markers had been prove to be versatile and the informativeness of these polymorphic markers depends upon the number of alleles and their relative population frequencies (Payne, 1997).

According to the Ruitberg *et al* (2000) it were used to generate a nationwide DNA database called the FBI Combined DNA Index System (CODIS) that are useful to support the crime scene evidence.

1.2 *Orang Asli* populations

Orang Asli or the aborigines are the original peoples of the Peninsular Malaysia (Figure 1.1). According to the Center for *Orang Asli* Concern (www.coac.org.my), the *Orang Asli* numbering over 147,412 in year 2003 (Table 1.1), they comprise of 18 distincts (Table 1.2) in their cultural-linguistic groups, that live in the scattered villages and camps from the coastal mangrove swamps to the rainforest of the central mountain range. They are separated into three main tribal groups namely, Semang also known as Negrito, Senoi and Proto Malay. The *Orang Asli* are not homogenous. They have their own linguistic and geographical groups (Thangaraj *et al.*, 2006) and also varies in their phenotype and traditional economy. They make their living by some combination of hunting, gathering, fishing, farming, arboriculture, selling forest products, producing petty commodities and wage labor.

Orang Asli groups were politically autonomous until the 1950's. The British colonialist had placed them under the jurisdiction of a federal Jabatan Hal Ehwal Orang Asli in order to win them away from the influence of communist guerrillas operating from forest bases (Baer, 1999). Laws establish to control *Orang Asli* during this insurrection which officially ended in 1960's, are still in force today and the Jabatan Hal Ehwal Orang Asli continue to administer them.

The three main groups of *Orang Asli* are phenotypically and genotypically different. For instance, Semang or Negrito is phenotypically similar to the Africans and Southeast Asians (Thangaraj *et al.*, 2006) however their origin still remain a mystery. On the other hand, the Proto Malay also known as *Melayu Asli* or *Melayu Purba* in the Malay language is an ethnic group in Malaysia. Anthropologists traced this group of newcomers Proto Malay seafarers migrated from Yunnan to Malaysia, during the stone age period (www.sabrizain.org/malaya/malays). Phenotypically, the Proto Malay has a smaller body size, they were polite and shy. Senoi are the hunting and gathering people who were reported to make extensive use of lucid dreaming to ensure happiness and mental health. The word of "Senoi" means human being or person in the language of the aboriginal people who still managed to practice their traditional way of life and Senoi actually related to the two groups, the Temiar and the Semai (Domhoff, 2003).

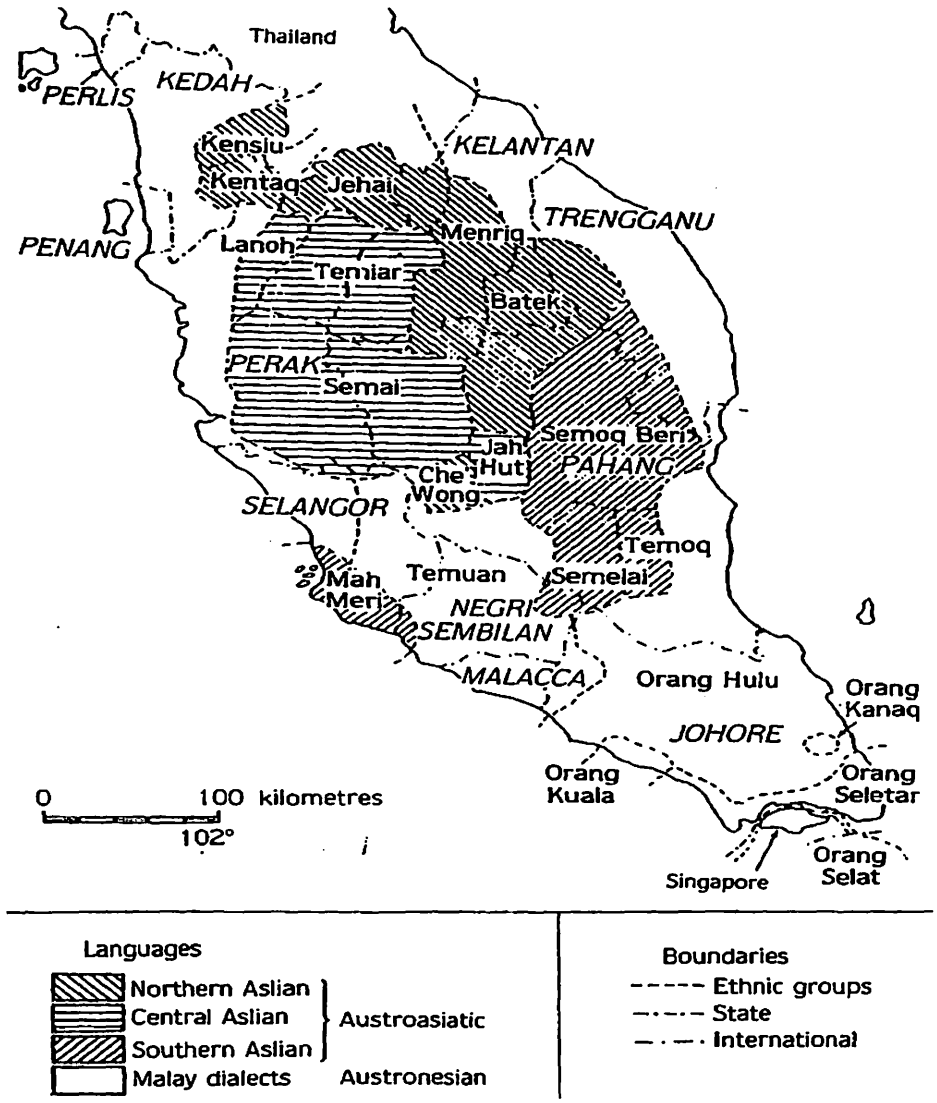


Figure 1.1: Map of Peninsular Malaysia indicating the distribution of the *Orang Asli* population

Table 1.1: Total of *Orang Asli* populations in Peninsular Malaysia. Data was collected from www.coac.org.my

Year	Number of individual
1960	43886
1965	45985
1969	52943
1974	56927
1989	72039
1993	92969
1998	116119
2000	133775
2003	147412

Table 1.2: *Orang Asli* population breakdown at year 2000. Data was collected from www.coac.org.my

Orang Asli Ethnic	Ethnic subgroup	Num.of individuals
Negrito	Kensiu	254
	Kintak	150
	Jahai	1244
	Lanoh	173
	Mendriq	167
	Bateq	1519
Senoi	Semai	34248
	Temiar	17706
	Jah Hut	2594
	Che Wong	234
	Mah Meri	3503
	Sema Beri	2348
Proto Malay	Temuan	18580
	Semelai	5026
	Jakun	21484
	Orang Kanaq	73
	Orang Kuala	3221
	Orang Seletar	1037
Total		113451

1.2.1 Negrito population

According to Thangaraj *et al* (2006), the Negrito population is believed to originate from the Andaman and Nicobar Islands which divided into the four sub groups such as Jarawa, Onge, Sentinelense and Great Andamanese. They were believed as the first aborigines to arrive in Peninsular Malaysia more than 10,000 years ago. The Negrito population is of interest in this study because of their unique genotypes obtained from the comprehensive molecular study using Y-chromosome and mitochondrial DNA (Thangaraj *et al.*, 2006).

The Negritos in Peninsular Malaysia are divided into six sub-tribes namely; Kensiu, Bateq, Mendriq, Jahai, Lanoh and Kintak. They normally live in the houses that made from bamboos and atap. A number of them stay in the houses provided by the government nowdays. Three of the tribes are located in Kelantan namely Jahai, Mendriq and Bateq.

1.2.1.1 Jahai

Jahai population is located at the Banun, Tiang River and near the Temenggor Dam in Perak. However, a number of them inhabit in Kelantan at Sungai Rual, Jeli and Hulu Kelantan (www.jheoa.gov.my). Phenotypically, Jahai populations are likely the same with Habsyi population or Negro population in Africa, tribes of Negrito in Andaman Island and Aeta The Philippines (www.jheoa.gov.my). Figure 1.2 shows the phenotype of the Jahai. Their houses are created like bunch of banana and bulid from bamboo and roofted by the 'bertam' and 'tepus'.

This population migrated from one place to another especially when one person has dead after disaster such as heavy rain and flood. On the other hand, they find new places for food and drink. Basically, the language of the Jahai referred to by the same name, is a member of the Northern Aslian subgroup of the Aslian languages, a branch of the Mon-Khmer language family. After the modernization, the Jahai language is heavily influenced by Malay, the Austronesian majority language of the peninsular (Burenhult, 2005).

Figure 1.2 Photo showing Jahai children in the bamboo house.
(Source: Health, Disease and Survival; a Biomedical and Genetic Analysis of the *Orang Asli* of Malaysia)

1.2.1.2 Mendriq

The Mendriq population is located upstream of the Kuala Lah river, Gua Musang district in Kelantan. Like Jahai population, the phenotypically appearance of Mendriqs are same like the Negrito population in Andaman island, Philippine and Southern Thai. Their bodies are small but tougher. This population was cited from Paleolithic age and their language is Austroassiatic. They involved in pure spoken of Mon-Khmer language but as civilization grow and their language was added with 'Bahasa' pronounciation (www.jheoa.gov.my).

1.2.1.3 Bateq

The Bateqs live in Taman Negara Kuala Tahan, around upstream of Tembeling River, Kechau River, Teluk Gunung in Pahang, Gala River, Chiku River, Tako River, Lebir River and Aring River in Kelantan and also in Berua River in Besut, Terengganu (www.jheoa.gov.my). In past times, they moved with possible reasons but in recent days they choose to live in new and safer place provided by the government.

2.0 OBJECTIVES

i. To optimize the PCR condition for primer D19S220 for GeneScan analysis purpose.

ii. To examine the PCR condition optimized to the Mendriq population.

3.0 MATERIALS AND METHODS

3.1 MATERIALS

3.1.1 10X TBE Buffer

An amount of 108 g (0.9M) Tris base and 55 g Boric Acid (0.9M) was dissolved in 500 ml distilled water. Then, solution was stirred and heated until dissolved. After that, 9.3 g EDTA (0.025M) was added to the solution and stirred again until dissolved. pH of the solution was measured and adjusted to pH 8.0 - 8.3 and put into a 1 L bottle with lowery tighten and cover with aluminum foil before autoclaving it.

3.1.2 1X TBE Buffer

100 ml of 10X TBE Buffer is diluted in 900 ml dH_2O

3.1.3 Agarose

An amount of 1 g agarose with 25 ml of 1X TBE buffer was used to make a 4% agarose gel. For the electrophoresis of genomic DNA, 1% agarose gel was applied.

3.1.4 Microsatellite Primer

Primers (D19S220) were synthesized by First BASE Laboratories Sdn. Bhd. The forward primer was fluorescence labeled at the 5' end. The primers were diluted into standard 100 μM for further use. Table 3.1 list the primer sequences applied in this study as well as the repetition at this locus.

Table 3.1 Primer sequence and repetition of locus D19S220

Primer	Sequence (5' to 3')	Repeat sequence
D19S220	F: Hex - ATG TTC AGA AAG GCC ATG TCA TTT G R: TCC CTA ACG GAT ACA CAG CAA CAC	(CA)n

3.1.5 ReddyMix PCR Master Mix

ReddyMix PCR Master Mix contains all components required for PCR reaction including the dye and precipitant to facilitate gel loading except for primers and DNA template. Composition of PCR Reddymix Master Mix is shown in Table 3.2.

Table 3.2 Composition of PCR Reddy Master Mix (Thermo Scientific)

Reagent
1.25 units ThermoPrime plus DNA polymerase
75 mM TrisHCL
20mM $(NH_4)_2SO_4$
1.5mM $MgCl_2$
0.1% (v/v) Tween 20
0.2mM each of dATP, dCTP, dGTP and dTTP
Precipitant and red dye for electrophoresis

3.2 METHODS

3.2.1 Selection of marker

The loci with heterozygozity of more than 0.75 were selected from the Genethon Microsatellite Map (http/www.genelink.wustl.edu/genethon/frame). To validate the specificity, the primer sequences selected were performed with BLAST analysis. These loci were optimized for optimum PCR condition. D19S220 was optimized into one Mendriq sample successfully and subsequently being tested in this population later in this study.

3.2.2 Sample Collection

Volunteers from the three selected tribes were recruited from Jahai, Mendriq and Bateq sub-tribes. Consent was obtained from the volunteers who fulfill the inclusion criteria of the study. The inclusion criteria are:

i) Subject must be a healthy person and free from any systemic diseases

ii) Must be descended from the same tribes for at least three generations

iii) Unrelated individuals

A total of 62 subjects who fulfill the inclusion criteria were recruited in this study. Three ml blood were withdrawn of the volunteers and kept in the EDTA bottle. Blood samples were then stored at -80°C in Human Genome Centre (HGC) Universiti Sains Malaysia for further experiment. Figure 3.1 showed that blood was taken from a volunteer by a nurse.