DISTRIBUTION OF HLA ALLELES IN THE SEMANG AND SENOI ORANG ASLI POPULATIONS IN PENINSULAR MALAYSIA

TASNIM BINTI ABD RAZAK

UNIVERSITI SAINS MALAYSIA

2015

DISTRIBUTION OF HLA ALLELES IN THE SEMANG AND SENOI ORANG ASLI POPULATIONS IN PENINSULAR MALAYSIA

by

TASNIM BINTI ABD RAZAK

Thesis submitted in fulfillment of the requirements

for the degree of

Master of Science

August 2015

ACKNOWLEDGEMENTS

First and foremost, I am grateful to Allah, the almighty God for enabling me to complete this MSc journey. This thesis is the product of MSc research which might not have been possible to be written without the help and support of the kind people around me, to only some of whom it is possible to give particular mention here.

Above all, I would like to thank my husband, Muhammad Naim Sulaiman, for his personal support and sacrifices. His great patience at all time has made this three years of MSc journey achievable, for which my sheer expression of thanks may not suffice. The birth of my son and twin girls during this MSc journey made me put more effort in completing this thesis. My parents and siblings, as well as in-law family have given me their unequivocal support throughout. Special thanks to my parents who inspired me to pursue my study.

My deepest appreciation goes to my principal supervisor, Professor Norazmi Mohd Nor for his advice, encouragement and guidance throughout the study. I am also extremely indebted to my co-supervisors, Associate Professor Dr. Zafarina Zainuddin and Mr. Sundararajulu Panneerchelvam, for their consistent guidance in understanding the fundamentals of research. I would also like to express my gratitude to Dr. Edinur Hisham Atan and Ms. Allia Shahril, for their consultation on statistic. I would like to thank staffs of School of Health Sciences (especially Forensic Science Laboratory, Biology Molecule Laboratory and Human Identification Unit) as well as Institute for Research in Molecular Medicine (INFORMM) for providing laboratory

facilities. Special thanks to Associate Professor Dr. Khairul Osman from UKM who inspired me to pursue postgraduate study in USM.

This thesis would not have been possible without the collaboration and teamwork of our sampling team, which comprise of USM Kubang Kerian, UiTM Puncak Alam and UiTM Shah Alam, especially Ms. Rose Iszati Ismet, who coordinated the sampling trip. It was an honour having a dedicated team comprising of Mrs. Nur Firdaus Isa, Mr. Nur Waliyuddin Hanis Zainal Abidin, Mrs. Sharifah Nany Rahayu Karmilla Syed Hassan and Ms. Lydiatul Shima Ashaari during the sampling trip. I thank the medical staffs from UiTM Sungai Buloh for conducting the blood withdrawal.

I would like to acknowledge the Ministry of Higher Education, Malaysia for funding this research under the Long Term Research Grant Scheme (304/PPSK/6150115/U132) and providing scholarship (MyMaster) that covers my tuition fees worth RM10, 000. Not to forget the Department of Orang Asli Development (JAKOA) for giving us the permission to conduct this research on Orang Asli. I would also like to thank all participants in this study. Without them, I may not able to complete this study.

Last but not least, thank you so much to my postgraduate friends for their support, motivation and ideas: Mr. Wong Weng Kin, Mr. Foo Phiaw Chong, Mrs. Rahayu Abd Gani, Mrs. Fatariah Zakaria, Mrs. Norhida Ramli, Ms. Nordiana Abu Bakar, Mrs. Norliyana Aris, Mrs. Zunura'in, Mrs. Aina Almardhiyyah and fellow friends who directly and indirectly contribute in the making of this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS, ABBREVIATIONS AND ACRONYMNS

APC	antigen-presenting cells
bp	base pair
D	linkage disequilibrium coefficient
ddH_2O	Deionized distilled water
DNA	Deoxyribonucleic Acid
dNTP	Deoxynucleotide triphosphate
EDTA	Ethylenediaminetetraacetic acid
g	gram
GSSP	Group Specific Sequencing Primer
GVHD	Graft-versus-host disease
HLA	Human Leukocyte Antigen
hrs	hours
HVS	hypervariable segment
HWE	Hardy-Weinberg equilibrium
JAKOA	Department of Orang Asli Development
kb	kilo base
kDa	kilodalton
KYA	kilo/thousand years ago
LGM	Last Glacial Maximum
MHC	Major Histocompatibility complex
ml	mililiter
min	minutes
mtDNA	Mitochondrial DNA
N	Number of samples

Na_2EDTA	Ethylenediaminetetraacetic acid disodium salt dihydrate
ng	nanogram
NGS	next-generation sequencing
nm	nanometer
PCR	Polymerase chain reaction
rpm	revolution per minute
RFLP	Restriction fragment length polymorphism
SBT	Sequence based typing
SEA	South East Asia
sec	second
SSOP	Sequence specific oligonucleotide probe
SSP	Sequence specific primer
Taq	*Thermus aquaticus*
TBE	Tris-borate-ethylenediaminetetraacetic acid
TCR	T-cell receptor
µl	microliter
UV	ultra violet
V	volt
WHO	World Health Organization
x *g*	G force or relative centrifugal force
°C	degree Celsius

TABURAN ALEL HLA DALAM POPULASI ORANG ASLI SEMANG DAN SENOI DI SEMENANJUNG MALAYSIA

ABSTRAK

Beberapa gelombang penghijrahan manusia ke Semenanjung Malaysia menyebabkan terbentuknya pelbagai lapisan kumpulan etnik. Penduduk terawal yang mendiami Semenanjung Malaysia adalah populasi Orang Asli iaitu Semang, Senoi dan Melayu Proto. Dalam kajian ini, penjenisan HLA untuk lokus HLA-A, -B dan -DRB1 dijalankan ke atas dua populasi Orang Asli yang terawal iaitu Semang dan Senoi. Penjenisan HLA dengan mengunakan teknik jujukan DNA telah dijalankan ke atas 85 individu daripada empat subkumpulan Orang Asli: Kensiu (N = 21) dan Bateq (N = 16) dari kumpulan kaum Semang, dan Che Wong (N = 10) dan Semai (N = 38) dari kumpulan kaum Senoi. Sebanyak 11 alel HLA-A, 12 alel HLA-B dan 17 alel HLA-DRB1 telah dikenalpasti daripada subkumpulan Orang Asli yang dikaji. *HLA-A*02:01* merupakan alel yang paling kerap ditemui dalam Kensiu, Bateq dan Che Wong manakala *A*24:07* adalah alel yang paling kerap ditemui dalam Semai. *B*18:01* merupakan alel yang paling kerap ditemui dalam semua subkumpulan Orang Asli. *DRB1*09:01* merupakan alel yang paling kerap ditemui dalam Kensiu dan Che Wong manakala *DRBI*12:02* merupakan alel yang paling kerap ditemui dalam Bateq and Semai. Berikut merupakan haplotip dua-lokus yang kerap ditemui dalam subkumpulan Orang Asli: *HLA-A*02:01-B*18:01, A*02:01-DRB1*09:01, A*02:01-DRB1*12:02, B*15:13-DRB1*12:02* dan *B*18:01-DRB1*09:01*. *HLA-A*02:01-B*18:01-DRB1*09:01* adalah haplotip tiga-lokus yang ditemui dalam semua subkumpulan Orang Asli. Analisis HLA dalam kajian ini mendapati drif genetik berlaku dalam subkumpulan Orang Asli berdasarkan bukti diversiti alel yang

terhad. Analisis penghubungan (*correspondence analysis*) menggunakan frekuensi alel HLA-A, -B dan -DRB1 mendapati bahawa subkumpulan Orang Asli berkongsi pertalian genetik yang tinggi dengan populasi dari Selatan China, Indochina, Malaysia, Singapura dan Indonesia. Walaupun data HLA ini tidak mencukupi untuk mengaitkan Semang dengan populasi Afrika, terdapat seorang individu Kensiu mempunyai alel *DRB1*15:03*, di mana alel ini ditemui dalam kekerapan yang tinggi di Afrika. Ciri alel HLA Semang menunjukkan profil yang lebih cenderung kepada populasi Asia Tenggara disebabkan oleh tekanan evolusi ke atas HLA. Data HLA juga menunjukkan Senoi mempunyai profil alel HLA yang lebih kurang sama dengan populasi dari Indochina/China Selatan. Kesimpulannya, profil alel HLA subkumpulan Orang Asli mempunyai ciri-ciri yang serupa dengan populasi dari Asia Tenggara. Kewujudan pangkalan data HLA untuk beberapa subkumpulan Orang Asli yang dikaji boleh dijadikan sebagai sumber rujukan yang bermanfaat dalam aplikasi klinikal (contohnya pemindahan sel dan organ), kajian hubung kait HLA dengan penyakit, kajian hubung kait HLA dengan tindakbalas terhadap dadah/ubat serta sebagai penanda DNA untuk mengenalpasti individu dan genetik sesebuah populasi.

DISTRIBUTION OF HLA ALLELES IN THE SEMANG AND SENOI ORANG ASLI POPULATIONS IN PENINSULAR MALAYSIA

ABSTRACT

Several waves of human migration into Peninsular Malaysia led to the stratification of various ethnic groups. The earliest inhabitants of Peninsular Malaysia are the Orang Asli population, namely Semang, Senoi and Proto Malay. In the present study, we typed the HLA-A, -B and -DRB1 loci of the two oldest Orang Asli populations, Semang and Senoi. Sequence-based HLA typing was performed on 85 individuals from four Orang Asli subgroups: Kensiu (N = 21) and Bateq (N = 16) of the Semang major group, and Che Wong (N = 10) and Semai (N = 38) of the Senoi major group. A total of 11, 21 and 17 HLA-A, -B and -DRB1 alleles were identified, respectively in the studied Orang Asli subgroups. *HLA-A*02:01* was the most common allele in Kensiu, Bateq and Che Wong, and *A*24:07* was the most common in Semai. *B*18:01* was the most common allele in all of the Orang Asli subgroups. *DRB1*09:01* was the most common allele in Kensiu and Che Wong, and *DRBI*12:02* was the most common allele in Bateq and Semai. The following two-loci haplotypes were common in the Orang Asli subgroups: *HLA-A*02:01-B*18:01, A*02:01-DRB1*09:01, A*02:01-DRB1*12:02, B*15:13-DRB1*12:02* and *B*18:01-DRB1*09:01*. The three-loci, *HLA-A*02:01-B*18:01-DRB1*09:01* haplotype was observed in all the Orang Asli subgroups. The HLA analysis in this study suggested that the Orang Asli subgroups experienced genetic drift based on evidence of limited allelic diversity. Correspondence analysis based on HLA-A, -B and -DRB1 allele frequencies suggested that the Orang Asli subgroups share high genetic affinity with the populations from South China, Indochina, Malaysia, Singapore and Indonesia.

Although the HLA data is not conclusive to correlate the Semang with the African population, there was one individual who expressed *DRB1*15:03* allele in the Kensiu where this allele is observed in high frequency in Africa. The Semang HLA allelic characteristics showed similar profiles with those of other Southeast Asian populations probably due to evolutionary pressure on the HLA gene complex. The HLA data also demonstrated similar HLA allelic profiles of the Senoi with Indochina/South China populations. In conclusion, HLA allelic profiles of the Orang Asli subgroups can be said to be characteristically Southeast Asian. The establishment of HLA database for selected Orang Asli subgroups may provide useful reference in clinical applications (e.g. cell and organ transplants), HLA-disease association studies, HLA-adverse drug reaction association studies and DNA marker for personal identification and population genetics.

CHAPTER 1

INTRODUCTION

1.1 Orang Asli: The Indigenous Group of Peninsular Malaysia

Malaysia is situated in the heart of Southeast Asia (SEA). It is separated into two regions namely Peninsular Malaysia and East Malaysia. Malaysia consists of 13 states and three Federal Territories. According to the Department of Statistics Malaysia (2012), the population of Malaysia is 28.96 millions in 2011. Malaysia is a multi-ethnic country comprising Malays, Chinese, Indians, Orang Asli and other ethnic groups. Based on year 2010 census, the Orang Asli is a group of indigenous people of Peninsular Malaysia that make up approximately 0.6% of the total Malaysian population. The Orang Asli is divided into three main groups which are, Semang (Negrito), Senoi and Proto Malay for administrative purposes by the Department of Orang Asli Development (JAKOA) based on their physical characteristics, linguistic affinities and cultural practices. Each group is further divided into six subgroups as listed in Table 1.1. The distribution of Orang Asli in Peninsular Malaysia is shown in Figure 1.1.

Table 1.1 The Orang Asli groups and subgroups

Semang	**Senoi**	**Proto Malays**
Bateq	Che Wong	Jakun
Jahai	Jahut	Orang Kanaq
Kensiu	Mah Meri	Orang Kuala
Kintak	Semai	Orang Seletar
Lanoh	Semoq Beri	Semelai
Mendriq	Temiar	Temuan

Subgroups name are based on JAKOA website (http://www.jakoa.gov.my/orang-asli/suku-kaumbangsa/).

Figure 1.1 Map showing the approximate distribution of the Orang Asli subgroups in Peninsular Malaysia
This map was redrawn and modified from Benjamin (2012).

Orang Asli is the Malay word for aborigines. ‘Orang’ mean people whereas ‘asli’ comes from Arabic word ‘asali’, means original, well-born or aristocratic. The Orang Asli was once called Sakai in the early British administration. In 1955, the Orang Asli was called ‘orang darat’. However, the term does not suit them because not all Orang Asli live in deep jungle. There are some who live along the coastal area and work as fishermen (Carey, 1976). Today, the government provides houses for Orang Asli to live comfortably and the permanent settlements are supplied with electricity and clean water.

According to the Federal Constitution (1963), the Orang Asli is defined as an aborigine of the Peninsular Malaysia. The Aboriginal Peoples Act 1954 (Revised 1974) further described the Orang Asli as any individual who is a member of an aboriginal ethnic group, who speaks an aboriginal language, and habitually follows an aboriginal lifestyle, customs and beliefs. The definition of Orang Asli also applied to adopted non-Orang Asli children and offspring of Orang Asli women whom married to non-Orang Asli men as long as they satisfy the above conditions. In other words, an Orang Asli is defined more by their culture instead of genetic traits.

In the present study, only the Semang and Senoi were included for Human Leukocyte Antigen (HLA) typing (see Section 1.2 for details on HLA). The Proto Malays were excluded because this group seems to be heterogeneous with respect to culture, religion and way of life (Carey, 1976) and are similar to the Deutro-Malays in terms of morphology, culture and language (Kasimin, 1991). Thus, it would be difficult to discuss them as a single group.

1.1.1 Semang (Negrito)

The Negrito in Peninsular Malaysia is also known as Semang. The word 'Negrito' means 'little negro'. They are known by their distinct physical characteristics: short, dark skin, broad nose and frizzy hair similar to the Africans. They consist of six subgroups, which are the Kintak, Kensiu, Bateq, Mendriq, Jahai, and Lanoh. (Carey, 1976). They used to be nomadic and live deep in the jungle (Carey, 1976). Nowadays, the Semang reside in permanent settlements in the central, northern and eastern regions of Peninsular Malaysia (Nicholas, 2006).

The Semang speak the Northern Aslian language belonging to the Mon-Khmer branch of the Austroasiatic language. Exceptionally, the Lanoh people speak the Central Aslian language (Benjamin, 1983). There is no evidence of Semang spoke another language previously other than Aslian of Austroasiatic language stock (Benjamin, 2013). The fact that both Semang and Senoi speak Austroasiatic language could be due to language shift which could have occurred in the Semang as a result of migration from Senoi's ancestors that introduced the Aslian language to Semang (Bellwood, 1993).

The Semang is the smallest group, accounting for approximately 3% of the total Orang Asli population. Interestingly, the Semang is believed to be the earliest group to inhabit Peninsular Malaysia, about 60,000 years ago. The Semang was thought to come from Africa and migrated throughout SEA before colonizing the Australasian region (Hill *et al.*, 2006). They are the present day descendants of the early Hoabinhians, who were largely nomadic hunters (Carey, 1976).

1.1.2 Senoi

The Senoi is the largest group of the Orang Asli in Peninsular Malaysia, constituting about 55% of the total population. The word 'Senoi' means 'people' or 'mankind' which was taken from the Temiar language. They comprise the Mah Meri, Semoq Beri, Temiar, Che Wong, Jahut, and Semai. The Senoi's skin is much lighter than the Semang and their hair is wavy. They are not nomadic but practise shifting cultivation (Carey, 1976).

They inhabited Peninsular Malaysia during the second wave of migration from South Asia, the mountainous areas of Cambodia, Vietnam and Myanmar (Baer, 1999). The Senoi group speaks various languages of the Aslian sub-family. The Mah Meri and Semoq Beri are Southern Aslian speakers. The Temiar, Semai and Jahut are Central Aslian speakers. Interestingly, the Che Wong is Northern Aslian speaker similar to the Semang (Benjamin, 1983). The relationship of the Aslian to the Mon-Khmer languages reflects their ancient connection with mainland SEA (Nicholas, 2006). However, some believe the Senoi to be descendants of the Australoid from Australia and Veddoid from South India (Skeat and Blagden, 1906).

1.1.3 Origin of the Orang Asli

The origin and route of the Orang Asli's migration into Asia is elusive. The popular migration routes namely northern or southern coastal routes have been the subject of much debate among scholars. Modern humans first migrated via the northern route across the Sahara, out of Egypt to the Levant (Metspalu *et al.*, 2006). Later, modern

humans successfully left Africa as a single group via the southern coastal route by crossing the mouth of the Red Sea from Eritrea, to India and on to SEA and subsequently reached the isolated Sahul continent (Oppenheimer, 2009).

A southern coastal route was the likely path taken since the coastal area offered a resource-rich living environment. The suggestion of the migration of Orang Asli via a southern route of migration is reasonable, as recent mitochondrial DNA (mtDNA) studies on relict populations of Southeast Asia, the Andaman and Nicobar Islands also point to human dispersals via the southern exit route (Macaulay *et al.*, 2005; Thangaraj *et al.*, 2005). Various theories have been proposed on the migrations of the Orang Asli into Peninsular Malaysia based on historical, linguistic and archaeological evidence as well as genetic studies (Carey, 1976; Benjamin, 1983; Macaulay *et al.*, 2005; Hill *et al.*, 2006). Recently, new hypothesis namely "Early Train" hypothesis has been proposed based on mtDNA and autosomal DNA data (Jinam *et al.*, 2012). However to date, studies on the intra-relationships among the Orang Asli population are limited.

1.1.4 Previous Genetic Studies on the Orang Asli

Macaulay *et al.* (2005) and Hill *et al.* (2006) studied the mtDNA of 260 maternally unrelated Orang Asli individuals comprising the Semang, Senoi and Proto Malay groups. They concluded that the Orang Asli groups have experienced increased levels of genetic drift. Nevertheless, phylogeographic traces of the ancestry of their maternal lineage remain. Their findings also concluded the Semang as the oldest inhabitants settled in SEA at least 50,000 years ago.

Zainuddin and Goodwin (2004) studied the mtDNA of Jahai and Kensiu from the Semang. They reported that many individuals of Orang Asli shared the same haplotype thus reducing the power of discrimination. Therefore, the application of mtDNA of Orang Asli in forensic field is limited due to the small number of haplotypes found in the Semang subgroups.

Bekaert *et al*. (2006) compared mtDNA and Y chromosome in Malay and Orang Asli populations. They concluded that the mtDNA and Y chromosome of the Malay population in Peninsular Malaysia show high levels of diversity. They further added that both the Malay and Orang Asli populations show a similar level of divergence between the different Y chromosome haplotypes.

Ang *et al.* (2011) studied the non-coding region of mtDNA from HVS I region and the cytochrome *b* (Cyt *b*) gene in 18 subgroups of Orang Asli in Peninsular Malaysia. One individual was chosen representing each subgroup. The maximum parsimony analysis based on hypervariable segment I (HVS I) region showed that the Orang Asli are closely related and were clustered together in different clades. The close relationships of the subjects in Cyt *b* locus also suggested a single wave of migration and entry into SEA (HUGO *et al.*, 2009).

Determination of phylogenetic relationship was also done using classical genetic markers. Saha et al. (1995) studied glucose-6-phosphate-dehydrogenase (G6PD) and phosphogluconate dehydrogenase (PGD) loci of 349 individuals of the Semai. Their study revealed close relationship between the Semai and the Khmer of Cambodia. They further added that the Semai are more closely related to the

Javanese than the Malay, Chinese and Tamil Indians. Based on G6PD and PGD, there are no evidence for close genetic relationships between the Semai and the Veddah or other branch of the Austroasiatic language family as reported by Fix (2008).

Endom *et al.* (2013) had screened three African-specific markers from G6PD gene namely the PvuII Type 2 polymorphism and A-mutation; and sickle cell trait in Malaysians (Semang, Malay, Chinese and Indian) with G6PD deficiency. Based on their findings, all the studied Malaysian populations have the PvuII Type 2 polymorphism and thus supported the African origin. However, A- and sickle cell mutations were not detected in their samples. Their finding indicated that the PvuII Type 2 polymorphism is conserved and inherited from their African ancestor most likely because of its survival advantage against malaria parasite in their endemic region. Other than mtDNA, the PvuII Type 2 polymorphism is also well suited for tracing the African origin as well as malaria historical record in other world populations.

Hirayama *et al.* (1996) studied the HLA-B gene of 56 unrelated individuals of Orang Asli subgroups (Temuan, Semai and Temiar) using sequence-specific oligonucleotide probe (SSOP) method. Their study revealed that *HLA-B*15:13* appears frequently in the Orang Asli subgroups. *HLA-B*15:13* is thought to confer some level of resistance against severe malaria because this allele shares the C-terminal peptide binding pocket (pocket F) with an African resistant type against severe malaria, *HLA-B*53:01* (Hill *et al.*, 1991). This observation suggests that

malaria infection might have independently enhanced the selection of functional change in polymorphic portion of HLA-B gene in Africa and in Southeast Asia.

Jinam *et al.* (2010) first reported the high resolution HLA typing in 74 Orang Asli participants representing the Semang (Jahai and Kensiu) and Proto Malay (Temuan). The participants were typed for HLA-A, -B, -DRB1 and -DQB1 genes using sequence-based typing (SBT). Principal component analysis (PCA) based on HLA-A, -B and -DRB1 allele frequencies showed the Orang Asli has close affinities with other Southeast Asian populations.

1.2 Human Leukocyte Antigen (HLA)

The study of HLA started more than a century ago when biologist interested in cancer began to study the tumours that spontaneously develop in domesticated mice. Hoping to extend the duration of their investigations beyond the lifetime of a single mouse, a healthy mice was transplanted with tumours from the sick mice. However, the transplanted tumour was rejected by the healthy mice. Later, the mechanism was known to be due to an immune response. Experiments continued by using inbred stock of mice. The transplantation was accomplished and propagation of tumours became feasible. Based on the observation, it was suggested that the acceptance and rejection of tumour grafts in the recipient was controlled by one or more genetic factors (Marsh *et al.*, 2000).

The same immunological rejection was also experienced during the transplantation of healthy tissues. Further studies of the phenomenon were carried

out using generations of highly inbred strains of mice. Results showed that tissues transplanted between mice of the same inbred strain were accepted whereas transplants between different strains were always rejected. The investigators finally concluded that there were several genetic loci that may contribute to tissue rejection which was subsequently named the major histocompatibility loci. The name was given because they determine tissue compatibility. The major histocompatibility complex (MHC) can be defined as cell-surface macromolecules that allow the immune system to distinguish foreign cells from self cells (Marsh *et al.*, 2000).

In the beginning, the MHC of mouse was defined by serological method. Later, the MHC of humans was discovered by using similar approach. Based on the study of mouse and human immune response, two classes of MHC were defined, which are MHC class I and MHC class II. In humans, the MHC is called Human Leukocyte Antigens (HLA) because these antigens were discovered on the surface of leukocytes. The short arm of chromosome 6 in the band 6p 21.3 is the location of the HLA genes which comprise about 4 Mb (The-MHC-sequencing-consortium, 1999).

1.2.1 Classes of HLA

The HLA is categorized into three classes which are HLA class I, HLA class II, and HLA class III (Figure 1.2). HLA class I region encodes HLA class I genes while HLA class II region encodes HLA class II genes. HLA class III region is located in between HLA class I and II regions (Marsh *et al.*, 2000).

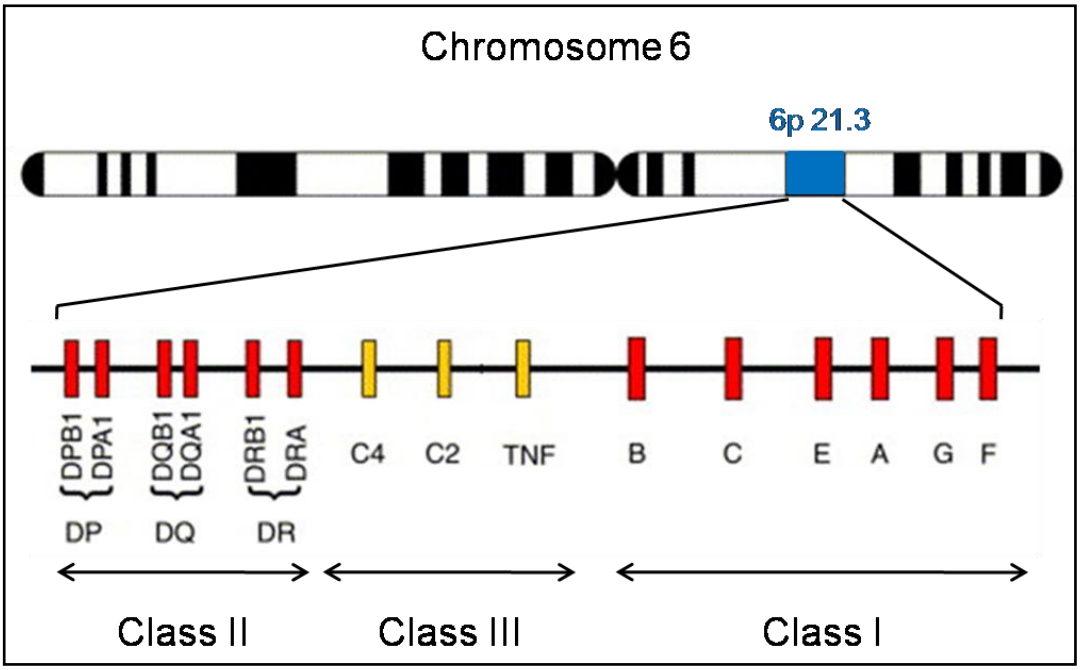

Figure 1.2 Simplified map of the HLA loci on the short arm of chromosome 6 The genes encoding class I and II molecules are represented in red and class III is represented in yellow. This figure is modified from Undlien *et al.* (2001).

The HLA class I genes are further divided into classical and non-classical genes. The classical class I genes encodes the α-chain of the antigen-presenting HLA-B, -C and -A molecules which are located telomeric in the complex. These antigens are found on almost every nucleated cell of the body except the central nervous system, skeletal and smooth muscle cells, parathyroid cells, pancreatic cells and corneal epithelium. Apart from that, the classical class I antigens are absolutely devoid in both men and women germinal cells. HLA-E, -G and -F are the non-classical class I genes and are found on placenta and extra villous membranes (Bodmer, 1987; Agrawal *et al.*, 2007). Apart from that, there are non-functional pseudogenes closely related in nucleotide sequence to the functional class I gene which are HLA-L, -J, -K, -H, -P, and -V (Marsh *et al.*, 2000; Robinson *et al.*, 2013).

The centromeric HLA class II genes encodes for HLA-DP, -DQ, and -DR genes. Class II expression is mainly limited to the antigen-presenting cells (APC) including B lymphocytes, macrophages, dendritic cells, and Langerhans cells. The class II region consists of a series of sub-regions, each containing A and B genes encoding α and β chains respectively. The DR gene family consists of a single DRA gene and up to nine DRB genes. The DQ and DP families each have one expressed gene for α and β chains and additional unexpressed pseudogenes (Choo, 2007). Other less polymorphic class II genes include HLA-DO and -DM genes (Marsh *et al.*, 2010).

In between the class I and class II region is the HLA class III region. The HLA class III genes do not encode HLA molecules but they encode complement proteins (C2, C4, factor B), 21-hydrolase, tumour necrosis factors (TNFs) and others

(Choo, 2007). The complement proteins interact with antibody/antigen complexes and help to destroy them by proteolysis. Apart from that, complement proteins also kill infected cells by disrupting the integrity of cell membranes (Snustad and Simmons, 2006). The HLA class III genes are not involved in the immune response and will not be included in this study.

1.2.2 Structure of the HLA

The HLA class I molecule consists of two chains, a heavy chain α of 43 kDa which is associated non-covalently with β_2-microglobulin (β_2-m) (light chain) of 12 kDa as shown in Figure 1.3 (Lechler and Warrens, 2000). The exon-intron organization of HLA class I is shown in Figure 1.4. Exon 1 encodes the leader peptide. Exons 2, 3 and 4 encode the three extracellular domains α_1, α_2 and α_3 respectively. Exon 5 encodes the transmembrane anchor while exons 6 and 7 encode the cytoplasmic tail. Exon 8 encodes the 3' untranslated region. The gene encoding β_2-m is located on human chromosome 15. The β_2-m gene is not given the HLA designation because of the chromosomal location. In addition, β_2-m is monomorphic and no polymorphic variants has been discovered (Marsh *et al.*, 2000).

HLA class II molecules are heterodimers consist of α and β chains of similar size (Figure 1.3). The α chains are of ~33-35 kDa and the β chains are of ~26-28 kDa. Both the α and β chains are encoded by genes from the HLA class II region. The exon-intron organization of HLA class II is similar to that of HLA class I as shown in Figure 1.5. Exon 1, 2 and 3 encode the leader peptide and domain chains for both α and β chains respectively. Transmembrane and cytoplasmic tail in α chains

are encoded by exon 4. The 3' untranslated region in α chains is encoded by exon 5. In β chains, the transmembrane region, cytoplasmic tail and 3' untranslated region are encoded by exon 4, 5 and 6 respectively (Marsh *et al.*, 2000).

1.2.3 Function of HLA

The HLA molecules perform a crucial role in regulation of the immune response through interaction with antigenic peptide, the T-cell receptor (TCR), and a number of costimulatory molecules (e.g. CD80, CD86) and adhesion molecules such as ICAM-1(CD54) and LFA-3 (CD58) (Bodmer, 1987; Navarrete, 2005). Endogenous and exogenous antigens presentation pathways are the two pathways involved in regulating immune response.

The HLA class I molecules are primarily involved in the presentation of endogenous antigenic peptides to $CD8^+$ cytotoxic T cells. The immune response is particularly important for recognition and destruction of virus-infected cells and early destruction of any cell that has produced aberrant self proteins. The presentation of exogenous antigenic peptides to $CD4^+$ helper T cells involves the HLA class II molecules. In this system, foreign proteins, microorganisms, or other antigens are taken up by antigen presenting cells and processed into peptides within cytoplasmic endosomes (see Rodey, 2000 for further details).

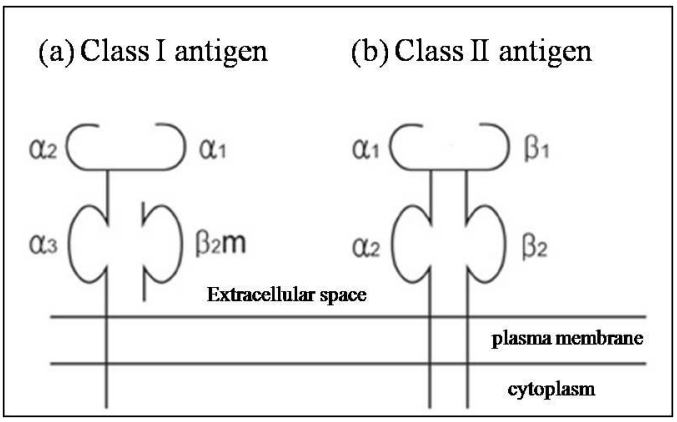

Figure 1.3 Schematic diagram of HLA (a) Class I and (b) Class II molecules
This figure is modified from Choo (2007).

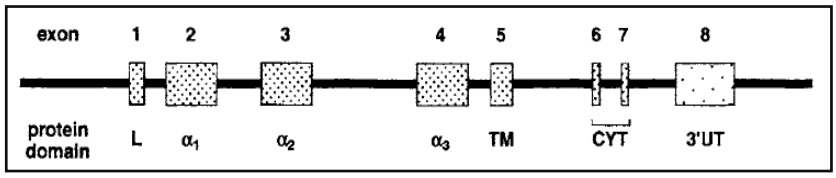

Figure 1.4 Schematic diagram of exon-intron organization of an HLA class I gene
L is leader peptide, TM is transmembrane region, CYT is cytoplasmic tail and 3'UT is 3' untranslated region. This figure is reproduced after Marsh *et al.* (2000).

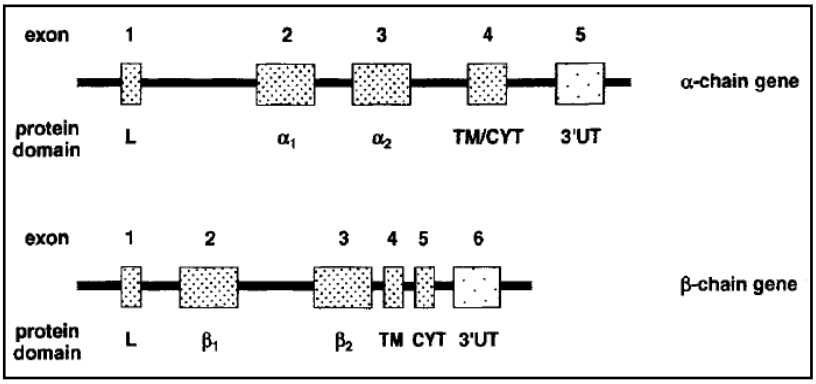

Figure 1.5 Schematic diagram of exon-intron organization of HLA class II
This figure is reproduced after Marsh *et al.* (2000)

1.2.4 HLA Alleles

According to the IMGT/HLA (ImMunoGeneTics/Human Leukocyte Antigen) database (http://hla.alleles.org/), 13, 369 HLA alleles have been described since July 2015. HLA-A, -B, and -C are highly polymorphic. The polymorphism is due to differences in the amino acid sequence of the HLA class I heavy chain. The differences arise from nucleotide substitutions in exons 2 and 3. As for HLA class II, polymorphisms are derived from both α and β chains. However, this depends upon the class II isoform. The genes encoding α chain are designated as A whereas the genes encoding for β chains are designated as B. In HLA-DR, α chain is monomorphic whereas β chain is polymorphic. HLA-DRB1 is the most polymorphic in class II gene (Marsh *et al.*, 2000). In HLA-DP and HLA-DQ, both α and β chains are highly polymorphic. The number of HLA alleles is shown in Table 1.2.

Table 1.2 Number of HLA alleles assigned as of July 2015

HLA	Gene	Alleles*
Class I	A	3, 192
	B	3, 977
	C	2, 740
	E	17
	F	22
	G	50
Class II	DRA	7
	DRB	1, 868
	DQA1	54
	DQB1	807
	DPA1	40
	DPB1	550
	DMA	7
	DMB	13
	DOA	12
	DOB	13
Total		13, 369

*As listed in http://hla.alleles.org/nomenclature/stats.html. HLA class I pseudogenes are not shown.

1.2.5 Nomenclature of HLA

World Health Organization (WHO) Nomenclature Committee for Factors of the HLA System is responsible for the naming of new HLA genes and allele sequences as well as their quality control. The committee first met in 1968 and meets regularly to discuss issues of nomenclature. The job of curating and maintaining a database of sequences has been the primary importance of WHO Nomenclature Committee. To date, they had published 19 major reports regarding the HLA antigens, genes and alleles. Collaboration between HLA Informatics Group and European Bioinformatics Institute has made it possible to disseminate new allele names and sequences to the public through web sites http://www.ebi.ac.uk/ipd/imgt/hla/ and http://hla.alleles.org (Marsh *et al.*, 2010).

In the beginning, the serologically defined leukocyte antigens were termed 'HL-A'. The term HLA replaced HL-A in 1975. Every single HLA allele name has an exclusive number, consisting of maximum 4 sets of digits, separated by colons. The sequence of the allele and its nearest relative contribute to the length of the allele designation (Marsh *et al.*, 2010).

The first 2 digits describe the type (allele family) which often corresponds to the serological antigen carried by an allotype. The next 2 digits describe the subtypes where numbers being assigned in the order in which DNA sequences have been defined. Alleles whose numbers differ in the two sets of digits must differ in one or more nucleotide substitutions that change the amino acid sequence of the encoded protein. The third set of digits is used to distinguish alleles that differ only by

synonymous nucleotide substitutions (also called silent or non-coding substitutions) within the coding sequence. The fourth set of digits are used to distinguish alleles that only differ by sequence polymorphisms in the introns or in the 5' or 3' untranslated regions that flank the exons and introns (Marsh *et al.*, 2010). The explanation of HLA nomenclature is simplified as shown in Figure 1.6 and Table 1.3 for further understanding.

Expression status of an allele can be indicated by the addition of optional suffixes. Alleles that have been shown not to be expressed are given the suffix 'N' for 'Null' alleles. Those alleles which have been shown to be alternatively expressed may have the suffix 'L', 'S', 'C', 'A' or 'Q'. No allele happen to be named with the 'C' or 'A' suffixes since July 2014 (http://hla.alleles.org/nomenclature/naming.html). The meaning of each suffix is presented in Table 1.4.

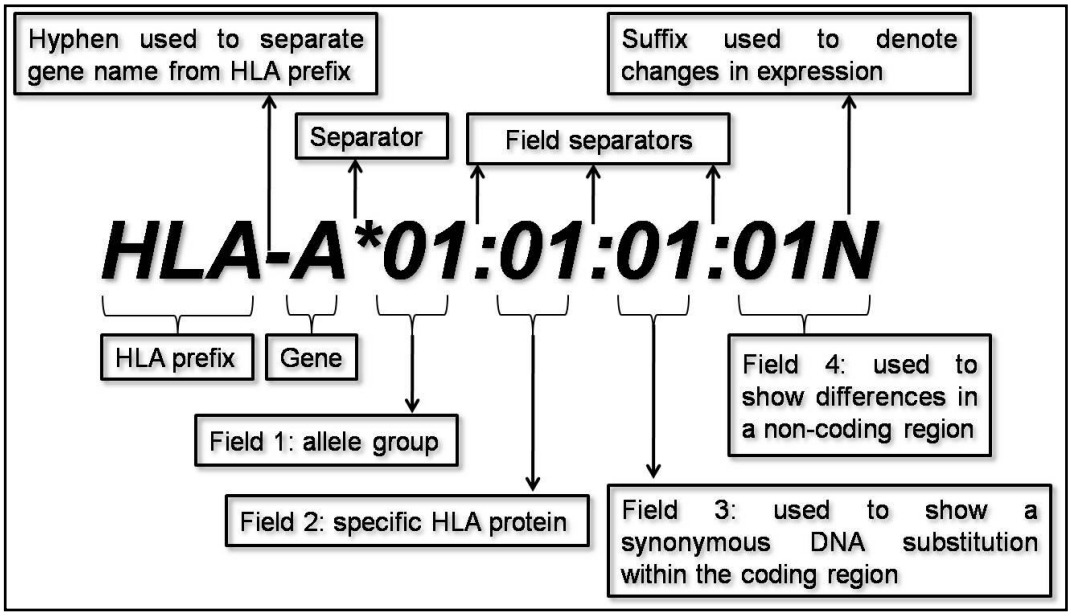

Figure 1.6 HLA allele nomenclature
This figure is modified from Marsh *et al.* (2010).

Table 1.3 HLA nomenclature

Nomenclature	**Indicates**
HLA	The HLA region and prefix for an HLA gene
HLA-DRB1	A particular HLA locus i.e. DRB1
*HLA-DRB1*13*	A group of alleles which encode the DR13 antigen or sequence homology to other *DRB1*13* alleles
*HLA-DRB1*13:01*	A specific allele
*HLA-DRB1*13:01:02*	An allele that differs by a synonymous mutation from *DRB1*13:01:01*
*HLA-DRB1*13:01:01:02*	An allele which contains a mutation outside the coding region from *DRB1*13:01:01:01*
*HLA-A*24:09N*	A 'Null' allele, an allele which is not expressed
*HLA-A*24:02:01:02L*	An allele encoding a protein with significantly reduced or 'Low' cell surface expression, where the mutation is found outside the coding region
*HLA-B*44:02:01:02S*	An allele encoding a protein which is expressed as a 'Secreted' molecule only
*HLA-A*32:11Q*	An allele which has a mutation that has previously been shown to have a significant effect on cell surface expression, but where this has not been confirmed and its expression remains 'Questionable'

This table was obtained from http://hla.alleles.org/nomenclature/naming.html.

Table 1.4 Optional suffixes for HLA nomenclature

Suffix	Meaning
N	A 'Null' or non-expressed allele
L	An allele has 'Low' cell surface expression when compared to normal levels.
S	An allele specifying a protein which is expressed as a soluble 'Secreted' molecule but is not present on the cell surface.
C	Indicate an allele product which is in the 'Cytoplasm' but not on the cell surface.
A	Indicate 'Aberrant' expression where there is some doubt as to whether a protein is expressed.
Q	Expression of an allele is 'Questionable' given that the mutation seen in the allele has previously been shown to affect normal expression levels.

This table was obtained from http://hla.alleles.org/nomenclature/naming.html.

1.2.6 HLA Typing Methods

Previously, HLA typing employed serological technique which was only able to distinguish limited polymorphisms. Nowadays, numerous alleles have been identified using the advance molecular approaches. The restriction fragment length polymorphism (RFLP) method was the first molecular technique used. However, RFLP typing requires a large amount of high molecular weight genomic DNA. Moreover, Southern blotting step in this technique is burdensome. This technique was not suitable for routine clinical typing. Most crucially, RFLP typing still unable to differentiate most of the HLA class II sequence polymorphism (Lechler and Warrens, 2000).

The study associated with genetic variation has revolutionized due to the emergence of Polymerase Chain Reaction (PCR) technology as well as sequencing techniques. PCR is used to amplify specific regions of an HLA gene using exon and locus-specific primers (oligonucleotides). These techniques include PCR-SSOP (sequence-specific oligonucleotide probes), PCR-SSP (sequence-specific primers), and sequence-based typing (SBT). The most extensive technique used for a complete HLA typing is SBT. The SBT technique can provide the highest resolution possible and is able to detect all available alleles. Most interestingly, this technique can also discover new alleles (Hoppe and Salama, 2007).

SBT method provides high-resolution typing, however, genotyping ‘ambiguity’ becomes a problem. Ambiguous genotyping has significantly reduced in the past few years with the launch of next-generation sequencing (NGS)