HLA GENETIC PROFILE OF MALAY SUB-ETHNIC GROUPS OF PENINSULAR MALAYSIA AS DETERMINED BY PCR-SSP AND SBT METHODS

ALLIA BINTI SHAHRIL

UNIVERSITI SAINS MALAYSIA

2016

HLA GENETIC PROFILE OF MALAY SUB-ETHNIC GROUPS OF PENINSULAR MALAYSIA AS DETERMINED BY PCR-SSP AND SBT METHODS

by

ALLIA BINTI SHAHRIL

Thesis submitted in fulfilment of the requirements

for the degree of

Doctor of Philosophy

August 2016

ACKNOWLEDGEMENT

In the name of Allah, the Beneficent, the Merciful.

I wish to express my sincere gratitude to Universiti Sains Malaysia (USM) Health Campus for providing me the opportunity to continue my studies here. I would also like to acknowledge the Ministry of Higher Education for giving me a scholarship thus enable me to pursue my dream.

My upmost gratitude goes to my supervisors, Professor Norazmi Mohd Nor, Dr. Zafarina Zainuddin and Mr. Panneerchelvam for their endless support, guidance and encouragement. Not to forget, Dr. Edinur Hisham Atan for helping me out especially in the process of writing and statistics.

I would also like to thank my colleagues, Puan Nur Haslindawaty, En. Nor Waliyuddin, Puan Tasnim, Puan Firdaus and Puan Sharifah Nany for helping me throughout my research work. I would also like to thank the assistant scientific officers Encik Azwan and Puan Rosniah for helping me during my laboratory work.

Last but not least, special thanks go to my beloved husband and my parents for their endless support and encouragement. They are my stimuli and without them I won't be able to be here. I love all of you so much.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Edinur et al., 2009, Dhaliwal et al., 2010, Jinam et al., 2010, Shi et al., 2010, Edinur et al., 2012). PCO map showing the first two principal components (44% cumulative variance). The solid triangles indicate the studied Malay sub-ethnic groups while solid circles indicate other populations.

LIST OF ABBREVIATIONS

AD	Anno Domini
APC	Antigen-Presenting Cell
AS-SEA	Austronesian of Southeast Asia
ATP	Adenosine Triphosphate
BC	Before Christ
Bp	Base pair
C2	Complement component 2
C4A	Complement component 4A
C4B	Complement component 4B
CD4	Cluster of Differentiation 4
CD8	Cluster of Differentiation 8
ddH_2O	Deionized distilled water
DNA	Deoxyribonucleic Acid
dNTP	Deoxynucleotide Triphosphate
EDTA	Ethylene diamine tetra acetic
EM	Expectation-Maximization
G	Gram
G	Guanine
HCl	Hydrochloric Acid
HLA	Human Leukocyte Antigen
HLA-SBT	Human Leukocyte Antigen Sequence-Based Typing
HLA-SSP	Human Leukocyte Antigen Sequence Specific Primer
Hsp70	70 kilodalton heat shock protein

HVS	Hypervariable Sequence
HWE	Hardy Weinberg Equilibrium
Kb	Kilo base
LD	Linkage Disequilibrium
$MgCl_2$	Magnesium Chloride
MHC	Major Histocompatibility Complex
MSY	Male specific region of Y chromosome
mtDNA	Mitochondrial DNA
MVSP3	Multivariate Statistical Package
N	Sample size
Na_2EDTA	Ethylene Diamine Tetra Acetic Acid Disodium
NaCl	Sodium Chloride
Ng	Nanogram
ng/µl	Nanogram per microliter
NJ	Neighbour Joining
NRY	Non-recombining region of Y chromosome
OD	Optical Density
Orange G	Orange Gelb
P	Probability of Random Match
PCO	Principle Coordinate Analysis
PCR	Polymerase Chain Reaction
PCR-SSOP	PCR–Sequence-Specific Oligonucleotide Probe
PCR-SSP	PCR–Sequence-Specific Primer
PHYLIP	Phylogeny Inference Package

RFLP	Restriction Fragment Length Polymorphism
RNA	Ribonucleic Acid
RPM	Revolutions Per Minute
rRNA	Ribosomal Ribonucleic Acid
SBT	Sequence Based Typing
SMM	Stepwise Mutation Model
SNP	Single Nucleotide Polymorphism
SSOP	Sequence Specific Oligonucleotide Probe
SSP	Sequence Specific Primer
STR	Short Tandem Repeat
T	Thymine
TAE	Tris Acetate Ethylene Diamine Tetra Acetic Acid
Taq	*Thermus aquaticus*
TBE	Tris Borate Ethylene Diamine Tetra Acetic Acid
TE	Tris EDTA
TNF	Tumor Necrosis Factor
Tris HCl	Tris Hydrochloric Acid
tRNA	Transfer ribonucleic acid
UEP	Unique Event Polymorphism
YAP	Y-chromosome Alu Polymorphism
β2	Beta-2

LIST OF PUBLICATIONS

Poster Presentations

Shahril Allia, Sundararajulu Panneerchelvam, Zainuddin Zafarina, Mohd-Nor Norazmi.Assessment of Genetic Relationship of Malay Sub-Ethnic Groups using HLA Class I and Class II Allele Frequencies. 2nd International Forensic Science Symposium, 14-16 November 2011, Le Meridien Hotel, Kuala Lumpur.

Shahril Allia, Sundararajulu Panneerchelvam, Zainuddin Zafarina, Mohd-Nor Norazmi. Human Leukocyte Antigen (HLA) Polymorphism in Four Malay Sub-ethnic Groups in Peninsular Malaysia. International Symposium on Forensic Science and Environmental Health 2009, organized by the Department of Chemistry Malaysia in Putra World Trade Centre (PWTC), 9-10th November 2009.

PROFIL GENETIK HLA DALAM KUMPULAN MELAYU SUB-ETNIK DI SEMENANJUNG MALAYSIA MELALUI KAEDAH PCR-SSP DAN SBT

ABSTRAK

Kepulauan Indo-Melayu merupakan tanah air bagi penduduk penutur Austronesia yang dikenali sebagai orang Melayu.Sekarang, definisi Melayu adalah terbatas kepada penduduk yang tinggal di Malaysia.Oleh kerana asal-usul orang Melayu di Malaysia yang berbeza dari pelbagai bahagian KepulauanIndo-Melayu, terdapat banyak sub-etnik Melayu yang tinggal di Malaysia.Objektif utama kajian ini adalah untuk mengkaji taburan alel HLA dan haplotaip dalam kumpulan sub-etnik Melayu.Taburan alel-alel antigen leukosit manusia (HLA) kelas I dan II telah disiasat pada 222 individu Melayu yang tidak mempunyai hubungan persaudaraan daripada enam kumpulan sub-etnik Melayu iaitu Kelantan (64), Champa (42), Patani (30), Kedah (30), Aceh (30) dan Mandailing (30) di Semenanjung Malaysia. Kumpulan alel HLA-A, -B, -C, -DQB1 dan -DRB1 lokus telah ditaip menggunakan kaedah primer khas jujukan (SSP). Analisa selanjutnya telah dijalankan dengan menggunakan kaedah pentaipan jujukan (SBT) ke atas 120 individu yang terdiri daripada empat kumpulan sub-etnik Melayu iaitu Kelantan (30), Champa (30), Patani (30) dan Mandailing (30) untuk menyiasat alel-alel bagi HLA -A, -B dan -DRB1. Bagi kaedah HLA-SSP, HLA-A*24, HLA-B*15, HLA-C*07, HLA-DQB1*03 dan HLA-DRB1*15 adalah kumpulan alel-alel yang paling kerap bagi gabungan sub-etnik Melayu (CMP) di lima lokus HLA yang telah disiasat.Haplotaip HLA yang paling kerap dipamerkan oleh CMP adalah HLA-A*02-C*08, HLA-C*08-B*15, HLA-DQB1*03-DRB1*12, HLA-A*02-C*15-B*08 dan HLA-A*33-B*44-DRB1*07. Bagi kaedah HLA-SBT, alel-alel HLA paling kerap di Kelantan, Champa,

Patani dan Mandailing masing-masing adalah; HLA-A*24:07, HLA-B*15:02 and HLA-DRB1*12:02; HLA-A*11:01, HLA-B*15:02 and HLA-DRB1*15:02; HLA-A*33:03, HLA-B*58:01 dan HLA-DRB1*12:02; HLA-A*02:01, HLA-B*15:13 dan HLA-DRB1*12:02. Hasil kajian menunjukkan bahawa terdapat persamaan dalam taburan alel HLA antara Kelantan, Champa dan Melayu Patani.Mandailing Melayu menunjukkan taburan HLA yang sedikit berbeza dan mempunyai hubungan genetik yang rapat dengan populasi Filipina berdasarkan kekerapan haplotaip HLA mereka. Secara keseluruhannya, kumpulan sub-etnik Melayu berkongsi hubungan genetik yang rapat di antara satu sama lain dan juga dengan populasi penutur Austronesia di Asia Tenggara. Kajian ini membentuk pangkalan data asas bagi penyelidikan ke atas populasi genetik, transplantasi dan juga kajian perhubungan HLA dengan penyakit.

HLA GENETIC PROFILE OF MALAY SUB-ETHNIC GROUPS OF PENINSULAR MALAYSIA AS DETERMINED BY PCR-SSP AND SBT METHODS

ABSTRACT

The Indo-Malay Archipelago is the homeland of the Austronesian speaking population known as the Malays. Now, the definition of Malay is confine to the population residing in Malaysia. Due to the different origin of the Malays in Malaysia from various parts of the Malay Indo-Archipelogo, there are many Malay sub-ethnic groups living in Malaysia. The main objective of this research is to study the distribution of the HLA alleles and haplotypes in the Malay sub-ethnic groups. The human leukocyte antigen (HLA) class I and II allele distributions were investigated in 222 unrelated Malay individuals belonging to six Malay sub-ethnic groups namely Kelantan (64), Champa (42), Patani (30), Kedah (30), Aceh (30) and Mandailing (30) in Peninsular Malaysia. The HLA-A, -B, -C, -DQB1 and -DRB1 allele groups were typed using sequence specific primer (SSP) method. Further analysis was carried out using sequence-based typing (SBT) method on 120 individuals comprising of four Malay sub-ethnic groups namely Kelantan (30), Champa (30), Patani (30) and Mandailing (30) Malays to investigate the HLA-A, -B and -DRB1 allele distribution. For the HLA-SSP method, HLA-A*24, HLA-B*15, HLA-C*07, HLA-DQB1*03 and HLA-DRB1*15 are the most common allele groups for the five HLA loci examined in the combined Malay population (CMP). The most common HLA haplotypes in the CMP are HLA-A*02-C*08, HLA-C*08-B*15, HLA-DQB1*03-DRB1*12, HLA-A*02-C*08-B*15 and HLA-A*33-B*44-DRB1*07. For the HLA-SBT method, the most common HLA alleles in Kelantan,

Champa, Patani and Mandailing respectively are; HLA-A*24:07, HLA-B*15:02 and HLA-DRB1*12:02; HLA-A*11:01, HLA-B*15:02 and HLA-DRB1*15:02; HLA-A*33:03, HLA-B*58:01 and HLA-DRB1*12:02; HLA-A*02:01, HLA-B*15:13 and HLA-DRB1*12:02. Based on the HLA distribution, Kelantan, Champa and Patani Malays bear similar HLA allele distribution.Mandailing Malays exhibits a slightly different pattern of HLA distribution and have close genetic relationship with the Filipinos based on their most frequent haplotypes. Overall, the Malay sub-ethnic groups share close genetic relationship with each other and also with other Austronesian speaking Southeast Asian population groups. This study forms a basic database for the research on population genetic, transplantation and disease association studies.

CHAPTER 1

INTRODUCTION

1.1 Background of the Study

The Malays, the Austronesian language speaking group inhabits the geographical region extending from the west of Indochina to the Bismark Island, known as Indo-Malay Archipelago (Bellwood, 2007). Besides these populations, the people of Cambodia and Vietnam from the Champ Empire and the people of south Thailand from the Patani Empire are also identified as Malays since they share the same culture and language family (Abdul Hamid, 2006, Porath, 2011). The Malays are the earlier dominant settlers of mainland Southeast Asia, but were later displaced by migrants from south and southwest China who later became the dominant settlers in countries such as Burma, Thailand, Vietnam, Laos and Cambodia (Bellwood, 2007).

Based on the Malaysian constitution, Malay is defined as a person who habitually speaks Malay, follows the Malay custom, professes Islam and whose ancestors are Malays. This is however, a very loose definition, that is challenging in any population investigation involving the Malays (Rashid, 2012). Therefore, for any population genetic studies to differentiate between these sub-ethnic groups, it is important to take into account their family background to avoid admixture in the general Malay population due to the existence of multi-ethnic groups in Malaysia.

The Malay population is descended mainly from the early Malay speaking groups who settled in the region and founded several ancient maritime trading states

and kingdoms. Due to trading activities, there were migrations of populations within the Indo-Malay Archipelago, which contributed to the presence of various Malay sub-ethnic groups in Malaysia (Bellwood, 2007). The convergence of multiple Malay sub-ethnic groups within Peninsular Malaysia provides an opportunity for population genetic studies of the Malays. This could give a more refine information on the genetic diversity within the Malay populations as a whole as well as among sub-ethnic groups. In addition to that, the genetic relationships among these Malay sub-ethnic groups could be studied and correlated with historical information.

Various genetic markers are employed in order to obtain information on the genetic diversity of a population. The genetic markers that are useful for population genetic studies include mitochondrial DNA, Y-Short Tandem Repeats (Y-STR), Short Tandem Repeats (STR), Single Nucleotide Polymorphism (SNP) and Human Leukocyte Antigen (HLA) (Cavalli-Sforza, 1998, Altukhov and Salmenkova, 2002, Sanchez-Mazas et al., 2014). For population genetic studies that serve to look into the relationship between populations or to trace the migration pattern of a population, neutral genetic markers (such as mitochondrial DNA (mtDNA), microsatellites or single nucleotide polymorphism (SNPs)) are mostly utilised in order to avoid other factors (selective) that could contribute to the change in the gene variants (Sunnucks, 2000, Altukhov and Salmenkova, 2002). However, in this study, we investigated the highly polymorphic HLA genetic marker in order to investigate the HLA genetic diversity of the studied populations and also the genetic affinity with other populations as this marker could provide information on the selection events involving the interaction between the individuals with their environment (Sommer, 2005).

The HLA consists of a group of linked genes located at the short arm of chromosome 6 (6p21.1-21.3) which plays an important role in immune response (Mehra, 2001). The role of HLA in antigen presentation to T lymphocytes is important in transplantation. In order to avoid organ rejection or graft vs host disease, the matching of HLA alleles between donor and recipient is necessary. HLA genes exhibit extensive polymorphisms that can cope with the myriad of potential pathogens in the environment (Turner, 2004). Therefore, due to its fundamental role in immune system, the HLA is under strong selection (Sanchez-Mazas et al., 2012). The distribution of HLA alleles and specific HLA haplotypes varies between different ethnic groups which reflect different selection pressures on populations evolving in geographically distinct environment. There are three factors that could shape a population based on their HLA genetic makeup; i) the distribution of allele frequencies, ii) the presence of ethnic specific HLA allele and iii) haplotypes that are linked together by linkage disequilibrium (LD) and are population-restricted (Turner, 2004). Linkage disequilibrium (LD) is defined as non-random association of alleles at two or more contiguous loci. Certain HLA alleles appear to be found together in a population more frequently than expected by random distribution based on their gene frequencies. This combination of linked HLA alleles is called haplotype. Certain combinations of HLA genes provide selective survival advantages in some populations which are revealed by the strong LD. Due to this, there is considerable conservation of combination of HLA genes on haplotypes that could reflect the original haplotype in which many others are derived (through recombination events) (Fernandez Vina et al., 2012). Due to the strong selection force that is acting on the HLA system, HLA allelic diversity should be analysed with regards to its haplotype and together with other genetic markers. Unlike mtDNA or SNPs genetic marker

which uses DNA sequence in the analysis of phylogeny, HLA uses allele frequency as an input data. HLA analysis does not produce a molecular clock that could place a time frame for the migration of a specific population. The reason for this is that the HLA system does not assume equal evolutionary rate across lineages as it is subjected to natural selection (Mueller, 2006, Fernandez Vina et al., 2012). In this study, we aimed at investigating the pattern of HLA diversity in the Malay population and also to see the genetic relationships between the Malay sub-ethnic groups and with other populations.

Due to the existence of a large number of HLA alleles and haplotypes present in the human population with varying frequencies, optimization of the recruitment of highly compatible donors in organ transplantation is very much needed. This can be done by increasing the number of donor pool size or an alternative approach is to refine the recruitment strategy by using the information of the patterns of HLA allelic diversity observed worldwide. This could facilitate the search for finding a HLA matched donor (Buhler et al., 2012). Besides transplantation, many HLA alleles are associated with particular diseases, for example, HLA-B*27 in spondyloarthropathies and HLA-DQB1*0602 in narcolepsy (Trabace, 2000). The data in this study will also provide information that will be useful as a reference for disease association studies on the Malay population.

1.2 Rationale of the Study

HLA profiling has been carried out on many populations world-wide due to its importance in population genetic studies, transplantation as well as disease association studies (Inotai et al., 2015, Nakaoka and Inoue, 2015, Augusto et al., 2016, Constantinescu et al., 2016, Jazairi et al., 2016, Kitpoka et al., 2016, Ravazzi-Gauch et al., 2016, Schafer et al., 2016, Varzi et al., 2016, Wu et al., 2016). There have been many studies recorded on the distribution of HLA alleles in the Malay population (Bugawan et al., 1999, Mack et al., 2000, Dhaliwal et al., 2007, Edinur et al., 2009). However, studies on the combination of the HLA-A, -B, -C, -DRB1 and -DQB1 loci together as a unit (haplotype) as well as HLA-C locus is still scanty (Edinur et al., 2009). High resolution data on HLA-B of Malay individuals are also lacking. Therefore, this study aims to provide a comprehensive picture of HLA genetic variations among the Malay sub-ethnic groups and the relatedness with other populations.

1.3 Objectives of the Study

1.3.1 General Objective

HLA profiling of unrelated Malay individuals belonging to different Malay subethnic groups namely Kelantan, Patani, Champa, Mandailing, Kedah and Aceh, selected using strict sampling criteria of sample collection.

1.3.2 Specific Objectives

i. To study the distribution of HLA-A, -B, -C, -DQB1 and –DRB1 allele groups for SSP method and HLA-A, -B and –DRB1 alleles for SBT method.
ii. To study the genetic affinity between the Malay subethnic groups and with other populations using HLA allele distribution.
iii. To identify possible new HLA allele in HLA-A, -B and DRB1 loci in Malay individuals.

CHAPTER 2

LITERATURE REVIEW

2.1 The Malays

2.1.1 The Origin of Malay population

Malaysia is a country situated in Southeast Asian region. There are two separate parts of Malaysia namely Peninsular Malaysia and East Malaysia separated by the South China Sea. Malaysia comprises of thirteen states and three federal territories. The total landmass of Malaysia is 329,847 km^2 with coordinate of 2°30'N 112°30'E. According to the Department of Statistics, 2012, the total number of people living in Malaysia was 29,336,800 in which 91.9% were Malaysian citizens and 8.1% were not. The Malaysian citizens consist of many different ethnicities namely the Malays (54.8%), other Bumiputeras (12.9%), Chinese (24.2%), Indian (7.3%) and others (0.9%). Bumiputera meaning "Son of the soil" is a Malaysian term describing the Malays and the indigenous populations of Malaysia.

The Orang Asli or the aboriginal groups of Malaysia are the first settlers in Malay Peninsula before the arrival of the Deutero-Malays, the current dominant ethnic group in Malaysia (Hill et al., 2006). There are three major groups of Orang Asli, characterised by their physical features, languages and cultural practices. The Semang group is the earliest Orang Asli to arrive in Peninsular Malaysia around 25,000 years ago and they are the descendants of the early Hoabinhians (Benjamin and Chou, 2002, Ang et al., 2011). The Senoi group of Orang Asli were from the second wave of migration into Malay Peninsula from South Asia, mountainous area

of Burma, Vietnam and Cambodia around 8,000 years ago. Both Semang and Senoi speak Aslian languages belonging to the Austroasiatic family (Benjamin and Chou, 2002, Hill et al., 2006, Ang et al., 2011). The Proto-Malay group of Orang Asli arrived after the occupation of Senoi group and they were from the Mongoloid group which later colonized the Indo-Malaysian Archipelogo (Bellwood and Sanchez-Mazas, 2005, Bellwood, 2007).

The Proto-Malay belongs to the Austronesian language family and they are similar to Deutero-Malay morphologically, linguistically and culturally (Benjamin and Chou, 2002). The Deutero-Malay is a mixture of different races such as Indian, Siamese, Arab, Javanese, Sumatran and Chinese (Comas et al., 1998, Hatin et al., 2011). It was also suggested that the Deutero-Malays migrated from southern China 1500 years ago and intermarried with the Proto-Malays and migrants from outside the Peninsula and eventually give rise to the modern Malays (Fix, 1995, Hatin et al., 2011). The different waves of migration of the Orang Asli groups into the Malay Peninsula however, are debated by many scholars (Baer, 2000, Bulbeck, 2000, Benjamin and Chou, 2002). They believe that differences between the Orang Asli groups are due to local adaptation or evolution rather than external migration. There are many hypotheses attempting to explain the migration pattern of the Malay race into the Malay Peninsula, but the two most accepted hypotheses are the "Out of Taiwan" and the "Slow Boat" model (Oppenheimer and Richards, 2001, Bellwood and Sanchez-Mazas, 2005).

The "Out of Taiwan" or better known as the "Express Train" model is mainly based on linguistic studies of the Austronesian dialogue and archeological findings

(Diamond, 2000, Bellwood, 2007). The majority of the people of Indo-Malaysian archipelago speak languages belonging to the Austronesian speaking population group (Bellwood, 2007) except the Orang Asli. The Austronesian speakers are distributed all over the world namely Indonesia, Malaysia, the Philippines, Taiwan, interior Vietnam, Madagascar, and Oceania. Diamond (2000) and Bellwood, (2007) proposed that the initial Austronesian speakers should have been from Taiwan who crossed from mainland of China as a result of population pressures arising from developments in agriculture around 4500-4000 BC. Around 2500-1500 BC, there were migration of people from Taiwan into the Philippines and Indonesia and into Borneo and Moluccas. There was further migration heading west towards Java, settling in parts of Mainland Southeast Asia namely Vietnam and Peninsular Malaysia by 500 BC. The migration reached as far as Madagascar. There were also migrations eastwards into the Pacific, settling as far as Easter Island. Figure 1.1 shows the timing of human settlement in the Austronesian speaking countries through the "Out of Taiwan" model.

The "Slow Boat" model on the other hand is mainly supported by genetic information of the mitochondrial DNA and Y-chromosomes and also by archaeological findings (Oppenheimer and Richards, 2001). Based on this theory, it was suggested that the origin of Austronesian speaking population is within Southeast Asia near the Wallace line 13,000 to 17,000 years ago (Oppenheimer and Richards, 2001, Matisoo-Smith and Robins, 2004, Friedlaender et al., 2008). The expansion out of South East Asia was triggered by the submerging of the Sunda shelf at the end of the last ice age, 7,000 to 14,000 years ago. The expansion took place in two migratory pathways: north into the Philippines and Taiwan and east into the

Pacific. The expansion of population from Southeast Asia and the spreading of the Austronesian languages are parallel and therefore it was suggested that the origin of Austronesian languages occurred in Southeast Asia during the Pleistocene period and radiated out through Melanesia and into the Pacific around 6,000 years ago (Oppenheimer and Richards, 2001). Figure 1.2 shows the origin of Austronesian speaking populations according to the "Slow Boat" theory.

There are many other less widely accepted speculations and hypothetical assertions by scholars relating to the origin of the Malay people. Genetic analysis of the Malays in Peninsular Malaysia may therefore provide the chronology of the prehistoric migration of Malays into this region.

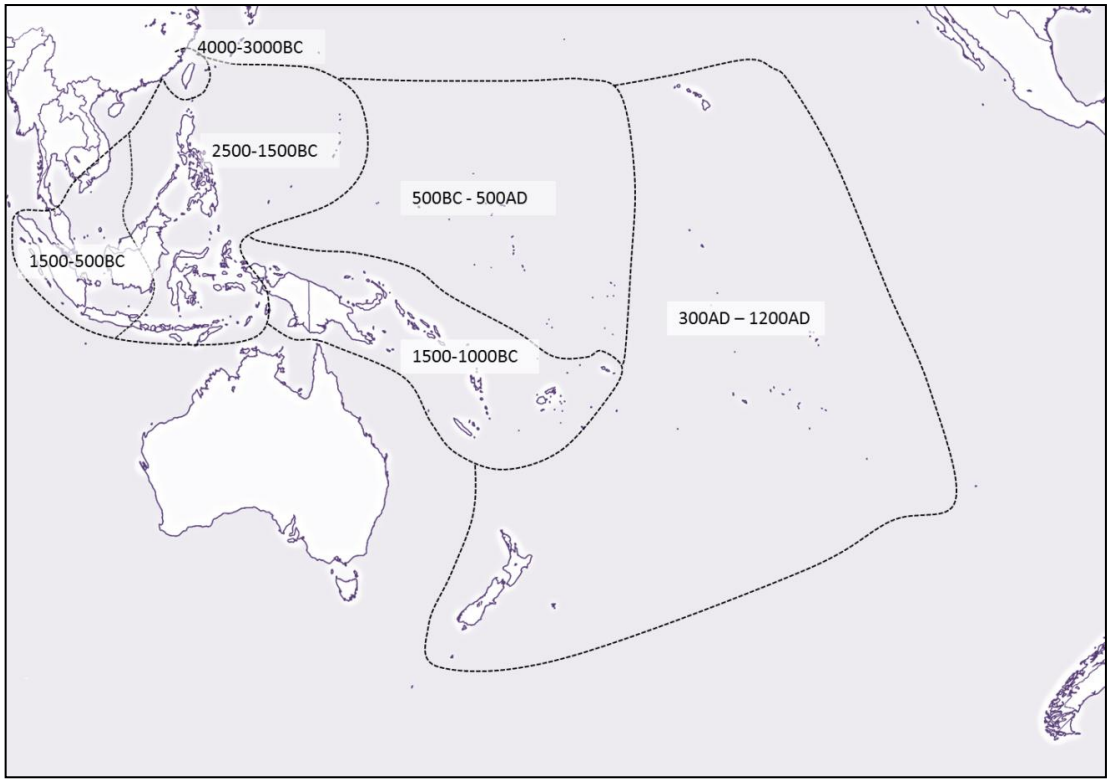

Figure 1.1: Map showing the approximate timing of human settlement in the Austronesian speaking countries through the "Out of Taiwan" model. The Austronesian speaking population originated in Taiwan in 4000 - 3000 BC. The expansion out of Taiwan ultimately reached the Malay Peninsula by 1500 - 500 BC.

Modified from: Bellwood,(2007) and http://commons:wikimedia.org/wiki/File:Oceania_full_blank_map.svg

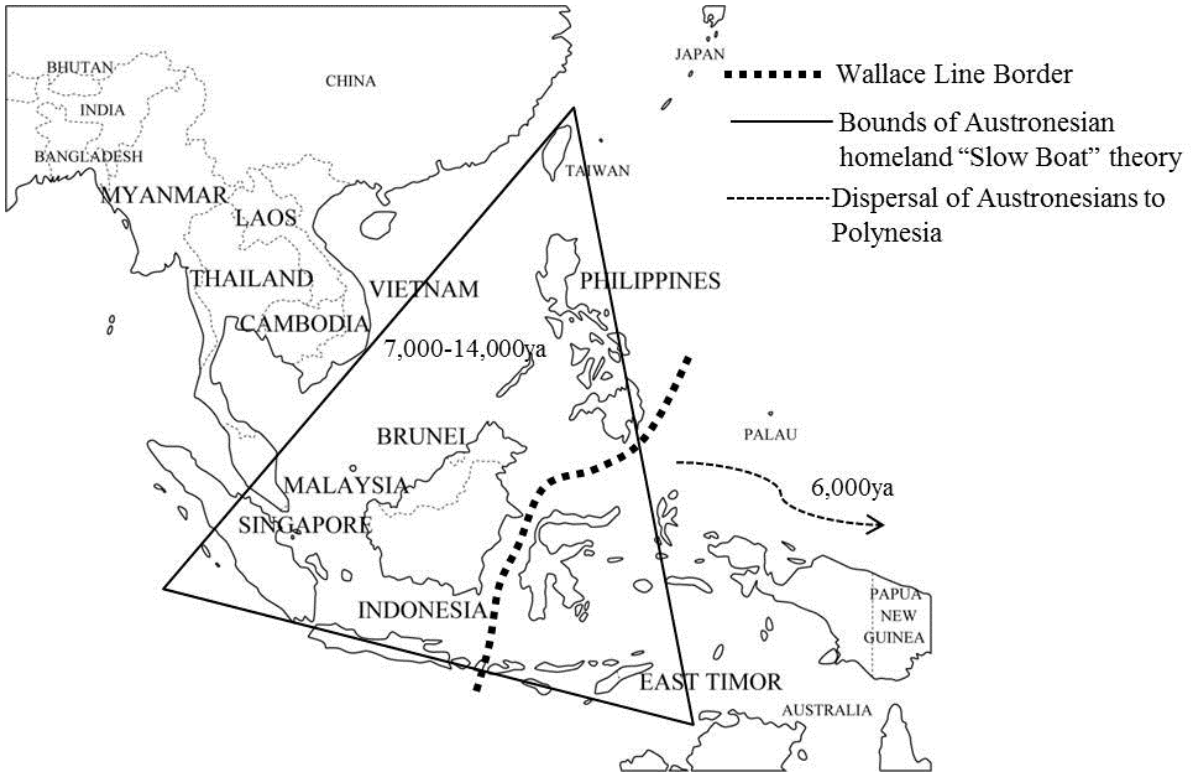

Figure 1.2: Map showing the origin of Austronesian speaking populations within the Southeast Asia around 13,000 to 17,000 years ago and migrated out around 7,000 to 14,000 years ago according to the "Slow Boat" theory.

Modified from:

http://english.freemap.jp/blankmap_dl.php?area=asia_e&country=continental_eastsouth&file_name=4.gif

2.1.2 Malay Sub-ethnic Groups

The Indo-Malaysian Archipelago which encompasses islands of Indonesia and Malaysia were initially occupied by series of groups of Malays belonging to the same culture and language family (Murphey, 2002). There are also groups of people from Cambodia and Vietnam who are descendants from the ancient Champa Empire and from South Thailand from the ancient Patani Empire who are considered as Malays because they share the same culture and language family (Halimi, 2008).

During the 19th century, there was migration of Malays from the Indonesian Islands to Peninsular Malaysia as a result of trading activities. The migration was further enhanced during the British colonization. As a result of this, the Malays inhabiting Peninsular Malaysia comprised of various sub-ethnic groups (Hasan, 1991).

In the Malay Peninsula, there are currently 15 Malay sub-ethnic groups (Aceh, Banjar, Batak, Bawean, Bugis, Champa, Jambi, Jawa, Riau, Kerinci, Mandailing, Minangkabau, Rawa, Yunnan and Patani) descendant from Indonesia, Thailand and Vietnam and six early Malay groups (Kedah, Kelantan, Perak, Terengganu, Pahang and Johor) of the Malay Peninsula. By studying the genetic distribution of each of the Malay sub-ethnic groups, the pattern of the genetic distribution unique to the Malay population could be identified, thus, enabling us to analyze the genetic affinity of the Malay population with other world populations. Edinur et al., (2009) had earlier studied the distribution of HLA alleles in six Malay sub-ethnic groups namely Kelantan,

Minangkabau, Rawa, Banjar, Bugis and Jawa Malays. In this study, we extended the study to assess the distribution of HLA-A, -B, -C, -DQB1 and -DRB1 in the Kelantan, Kedah, Mandailing, Aceh, Champa and Patani Malay sub-ethnic groups.

2.1.2(a) Kelantan

The Kelantan-Malays are known to have links with the Patani kingdom in the 18th century (Rahman, 1987, Rahman and Shukri, 2011). The geographic location of Kelantan provided opportunity for the Kelantan Malays to interact with populations in the northern part of Malay Peninsula (Hasan, 1991).

2.1.2(b) Patani

Patani is a district in Thailand located in close proximity with the Malaysia-Thailand border. Patani was once a strong Malay empire before it fell to the Siamese government in the 1785. As a result of this, there was a population exodus from Patani to Malaysia (Hasan, 1991).

2.1.2(c) Champa

Champa is the one of the earliest Malay civilization inhabiting Vietnam from the 2nd AD. The Champa people are Austronesian speakers. By the 1490s, Champa was absorbed by the Vietnamese government after its defeat. Many Champa people migrated out of their

original homeland to different countries such as Cambodia, Malaysia, Thailand and to many other places as far as the United States (Willoughby, 1999).

2.1.2(d) Kedah of Lembah Bujang

The Kedah-Malays from Lembah Bujang was also recruited as participants in this study. Lembah Bujang is one of the oldest and most historical sites in Malaysia. It is the site of Kedah's ancient civilization from 7th to 14th AD. Lembah Bujang was a popular trading center during the late 11th AD. The civilization was influenced by Hinduism/Buddhism and the presence of Chandi temples provides proof of this influence (Hasan, 1991). Based on historical records, Kedah was the first state in Malay Peninsula to become an Islamic state (Salleh and Salleh, 1998).

2.1.2(e) Aceh

Aceh is located in the northern part of Sumatra. Islam was first established in Aceh before it spread to other parts of Southeast Asia. In the beginning, migration of Aceh people to the Malay Peninsula was mainly for education and religious purposes. Trading was then carried out by the Aceh people in the Peninsula which ultimately resulted in the migration of Aceh traders to the Peninsula (Hing, 1972). During the Dutch invasion of Sumatra, many Aceh people fled to the Malay Peninsula, mainly to Kedah, since Kedah is nearer to Aceh. The Aceh people mainly settled in Kedah, Perak, Pulau Pinang and Pulau Langkawi (Hussin, 1980).

2.1.2(f) Mandailing

Mandailing people originated from the northern part of Sumatra Island in Indonesia. During the civil war in West Sumatera from 1821 until 1838, there was a massive migration of the Mandailing people to the West of Peninsular Malaysia. The migration of Mandailing people to the Malay Peninsula continued even after the civil war. They were then engaged mainly in mining, trading and mercenary activities as well as in politics. In Malaysia, they are found mainly in Selangor and Perak (Lubis, 2005).

2.1.3 Previous Genetic Studies on the Malay population

There are many population genetic studies carried out on the Malay population based on various genetic markers (Edinur et al., 2009, Hatin et al., 2011, Hatin et al., 2014, Loo and Gan, 2014, Norhalifah et al., 2015, NurWaliyuddin et al., 2014, Aghakhanian et al., 2015, Deng et al., 2015, Hoh et al., 2015, Manaf et al., 2015, Wan Syafawati et al., 2015).

Deng et al.,(2015) analysed 288,660 single nucleotide polymorphisms (SNPs) in 133 Malay individuals from Peninsular Malaysia and 50000 SNPs in 20 Minangkabau and 18 Kelantan Malays. This study showed that the Malays are composed of mixed ancestry namely East Asian, South Asian, Austronesian, and Southeast Asian aborigines around 175 to 1500 years ago. The Austronesian (Taiwan aborigines) and Southeast Asian aborigines (Proto-Malays) constitute largest proportion in the mixture. This study

suggested that geographical isolation and admixture are important factors in shaping the genetic structure and diversity of the Malays.

Aghakhanian et al. (2015) carried out a study on the Proto-Malays together with Negrito and Senoi Malaysian aboriginal groups using over 2,000,000 SNPs. Their study suggested that the Proto-Malays and Senoi are admixtures between Negritos and East Asian population. There were also evidence of gene flow between Austro-Asiatic-speaking aborigines and populations from Southeast Asians and South China populations. This indicates the presence of Austro-Asiatic-speaking aborigines in South East Asia even before the Austronesian expansion. This study supports the multiple waves of migration theory into Southeast Asia.

Hoh et al. (2015) analysed fine scale population sub-structure of 431 Malay individuals in Peninsular Malaysia and Singapore. Two clusters were observed; north and south Malay population clusters. This finding showed heterogeneity in sub-structuring of the Malays which may be due to the differences in their origin as well as geographical isolation.

Edinur et al. (2009) studied the low resolution Human Leukocyt Antigen (HLA) allele group distribution of six Malay sub-ethnic groups namely Kelantan (25), Minangkabau (34), Jawa (30), Bugis (31), Banjar (33), and Rawa (23). Based on the findings, the Malay sub-ethnic groups were believed to share close genetic relationship with each other and also with other Asian populations. However, the Banjar, Bugis and

Jawa and the Minangkabau and Rawa shared more similar HLA genetic make up compared to other Malay sub-ethnic groups. Kelantan Malays appear to be a more distinct group than the other Malay sub-ethnic groups.

Hatin et al. (2011) studied the genetic structure of Malay population by using 54,794 autosomal single nucleotide polymorphism (SNP) genotype data obtained from four Malay sub-ethnics comprising of 18 Kelantan individuals, 20 Minang individuals, 19 Jawa individuals and 14 Bugis individuals. Based on the findings, the Jawa, Bugis and Minang Malays showed close genetic relationship with each other while Kelantan Malays were more distinct than the other Malay subethnic groups and composed of a mixture of Indian and the Semang aboriginal group of Malaysia. Furthermore, Hatin et al. (2011) observed genetic structuring among the Malay populations which might be due to their different historical origins.

Loo and Gan (2014) studied the Kelantan Malays in the context of their historical, genetic and linguistic data. This study showed that the Kelantan Malays and the Semang (Jehai and Kensiu) are genetically similar and this is in agreement with linguistic data based on the fact that the Aslian language spoken by the Semang was probably transmitted through activities involving agriculture.

NurWaliyuddin et al. (2014) studied on the variation of KIR system on a total of 120 Malay individuals from four Malay sub-ethnic groups namely Kelantan, Jawa, Banjar, and Patani. Based on their findings, the Jawa, Patani and Banjar Malays are

genetically closer than the Kelantan Malays. The principal component analysis showed that the four Malay sub-ethnic groups clustered together with the other Southeast Asian populations. Overall, their study was in agreement with the common origin hypothesis of the Austronesian speaking population

Hatin et al. (2014) studied 54,794 SNP genotype data from a total of 472 individuals comprising of 5 Malay sub-ethnic groups namely Bugis, Jawa, Minang, Kedah and Kelantan, Thailand, Indonesia, China, India, Africa and the aborigines of Peninsular Malaysia. The findings of their study showed the presence of admixture in the Malay population with the Indians especially in Kedah and Kelantan. The Kedah and Kelantan Malays were genetically similar to the Thailand Patani Malays. The presence of admixture of the Malays with the Indians and Chinese were more evident in the Kedah, Kelantan and Thailand Patani Malays. Overall, this study showed that the Malays, Indonesians and Thailand Patani population shared common ancestry with the Proto-Malays and the Chinese.

Norhalifah et al. (2015) analyzed 22 SNPs in the cytokine genes in five Malay sub-ethnic groups namely Kelantan, Aceh, Mandailing, Minangkabau and Patani Malays. Based on their study, the Malay sub-ethnic groups were observed to share close genetic relationships with other Asian populations, Presence of admixture were evident in the Patani Malays with the Thailand population.

Manaf et al. (2015) and Wan Syafawati et al. (2015) studied the Human Neutrophil Antigen (HNA) and the Human Platelet Antigen (HPA) respectively on five Malay sub-ethnic groups namely Banjar (30), Bugis (37), Champa (51), Jawa (39) and Kelantan (35). Based on the findings of these studies, the Malay sub-ethnic groups were found to share close genetic relationships with other Asian populations and the genetic profile obtained from these studies are useful for future studies involving disease association.

2.2 Human Leukocyte Antigen (HLA)

Human leukocyte antigen (HLA) is the general term for a group of closely linked genes in the major histocompatibily complex (MHC) region located on chromosome 6 that encodes the cell- surface antigen-presenting proteins (Knapp, 2005). Figure 1.4 shows the map of the HLA region. HLA system was discovered by Jean Dausset in 1958 and the first to start HLA DNA typing. The DNA region of HLA spans around 3.6 megabases and is made up of multiple genes (Knapp, 2005). The main function of HLA is in immune response and its regulation. HLA is unique in each individual and is one of the most polymorphic immunological genetic systems in the human genome. MHC molecules consist of three major classes, class I, class II and class III (Wood, 2006). To date, there are 9749 HLA class I and 3274 class II genes (Gonzalez-Galarza et al., 2015).

2.2.1 HLA Classes

Class I HLA molecules are found on almost all nucleated cells. HLA class I molecules are composed of one heavy chain linked to β2 microglobulin (as shown in figure 1.3). The HLA class I region is composed of classical and non-classical class I genes. The classical class I genes includes HLA-A, -B and -C while the non-classical class I genes includes HLA-E, -F and -G. Classical class I genes are highly polymorphic compared to the non-classical class I genes which exhibit limited polymorphisms. Endogenous peptides bound to the peptide binding region of HLA class I molecules are presented to CD8-positive cytotoxic T lymphocytes activating the disruption of the target cell (Knapp, 2005, Wood, 2006). As to date, there are 3107 HLA-A allele, 3887 HLA-B allele and 2623 HLA-C alleles (Gonzalez-Galarza et al., 2015).

Class II HLA molecules consist of α and β chains linked together (as shown in figure 1.3). Class II HLA molecules are restricted to a few specialized antigen presenting cells (APC) which include dendritic cells, macrophages, B cells, activated T cells and cells that make up the interior of the thymus. Each of the three forms of class II, HLA-DR, -DQ and -DP, has two genes, A and B, encoding the α and β chains. Exogenous antigenic peptide binds to the HLA class II binding region of the APC. The CD4-positive helper T cells become activated when their T-cell receptor bind to the antigen peptide-HLA molecules and will stimulate the immune response (Turner, 2004). As to date, there are 1726 HLA-DRB1 alleles and 780 HLA-DQB1 alleles (Gonzalez-Galarza et al., 2015).

HLA class III composed of many genes with many functions. It is located in between class I and class II HLA loci. It contains genes for 21α- and 21β-hydroxylase genes, genes for the complement system such as C4A, C4B, C2 and factor B, Tumor necrosis factor (TNF) and also heat shock protein 70 (Hsp70) (Trabace, 2000). HLA class III has no structural or functional correlation with HLA class I and class II and exhibit lower polymorphism (Mehra, 2001).

The loci in the HLA complex are closely linked with each other. Therefore, the allele pair in each chromosome is contributed by; one allele from the mother and the other from the father to form a single genetic unit or haplotype, (Trabace, 2000). Figure 1.5 shows the pathways for HLA class I and class II for antigenic peptides.

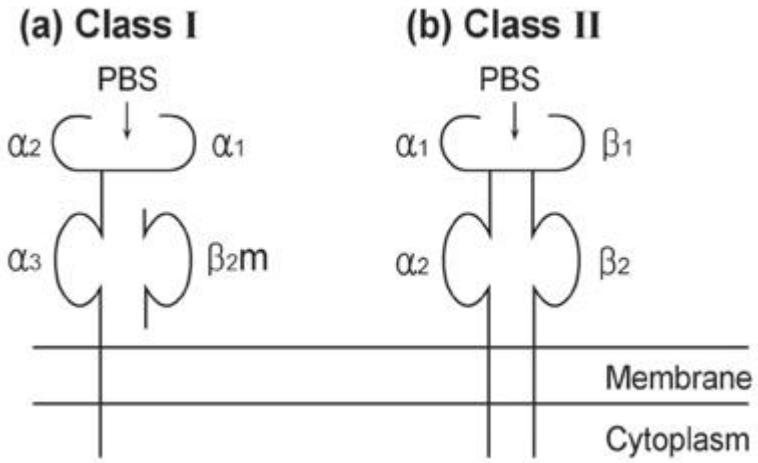

Figure 1.3: Structure of HLA class I (a) and class II (b) molecule from Choo (2007). a| HLA class I molecule. b| HLA class II molecule. Both class I and class II HLA molecules have a peptide binding groove (PBS) that will allow the binding of peptides for T-cell presentation.

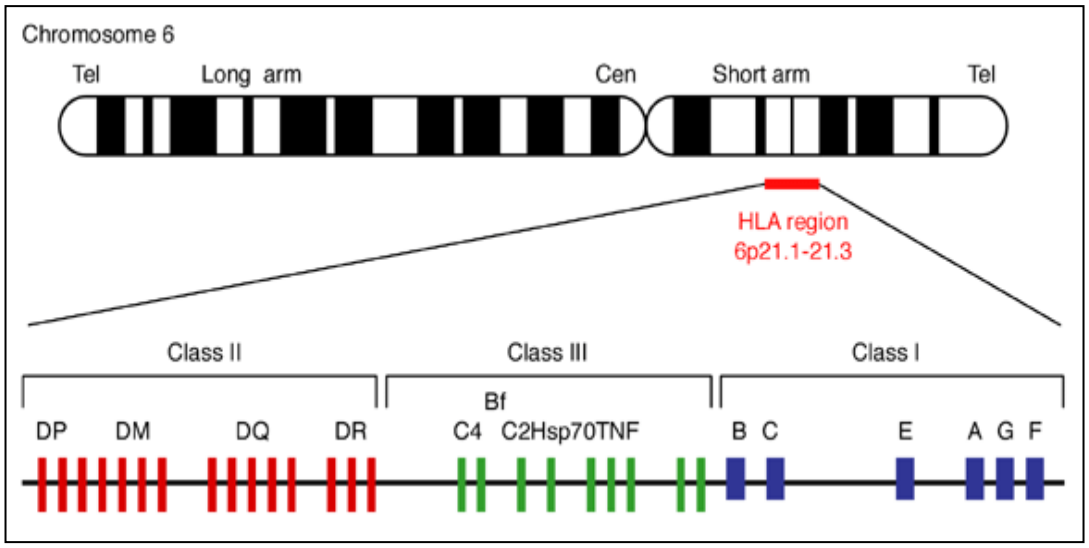

Figure 1.4: Map of the human leukocyte antigen (HLA) region from Mehra et al. (2010).

HLA class I and class II are shown in blue and red colored boxes respectively. The HLA genes are located on the short arm of chromosome 6 within the chromosomal band 6p 21.1 to 21.3. HLA class I and class II exhibit extensive polymorphism compared to HLA class III.