UNIVERSITI TEKNOLOGI MARA

WHOLE GENOME SEQUENCING AND ANALYSIS ON THE TRIOS OF THE CHE WONG AND SEMAI: UNDERSTANDING THE MEDICO-GENOMIC ASSOCIATION

ROSE ISZATI BINTI ISMET NAYAN

Thesis submitted in fulfilment
of the requirements for the degree of
Doctor of Philosophy

Faculty of Pharmacy

January 2017

ABSTRACT

As to date, there is no report on the architecture of the whole genomes of the unique Orang Asli subtribes in Peninsular Malaysia. The Che Wong was selected as the group of interest as they have a dwindling population of 651, while the Semai Orang Asli with a population number of 51,313 was selected as the control group. The Semai is the largest sub-tribe of the Senoi which is also the largest group amongst the Orang Asli. This study aims to understand the variation in genomics composition of the Che Wong and Semai Orang Asli by a deep and systematic characterization of the genome through development of an in-house bioinformatics pipeline. The disease risk and protection conferred by genetic traits were explored via sequencing the whole genome of a trio family of both the Che Wong and Semai sub-tribes. The Che Wong and Semai genomes were sequenced and mapped to the human reference genome (hg19) with an average coverage of 44.1x and 43.0x, respectively. A total of ~6.23 million and ~6.18 million variants were identified for the Che Wong and Semai trios, respectively. An average of 448,166 and 436,520 of these variants were found to be unique to the Che Wong and Semai trios, respectively. An average of 1.6% of the total variants called for the Che Wong and Semai genomes were known to be evolutionary conserved. A closer relationship between the Che Wong and Semai with the Asian populations than the African, American and European peoples was also observed. There is a distinctive difference in health status between the Che Wong and Semai trio where the Semai are seen to be in a healthier state. Medico-genomic association also revealed more disease impact for the Che Wong trio as compared to the Semai trio genomes. A total of 93 and 106 *de novo* mutations (DNMs) were identified for the Che Wong and Semai trio respectively. These mutations were also studied for their medico-genomic association. We report here for the first time the catalogues of the genomic architectures of the Orang Asli, the Che Wong and Semai in precise along with their medico-genomic findings. These data provide new perspectives of the genomics background of the indigenous populations in South East Asia (SEA) which we believe would be useful for the scientific and health community.

ACKNOWLEDGEMENT

First of all, I would like to praise my “syukur” to The Almighty Allah S.W.T. for giving me the strength and courage to persevere through my entire academic journey. Alhamdulillah, with His mercy and blessings; I had managed to complete this study: “Whole Genome Sequencing and Analysis on the Trios of the Che Wong and Semai: Understanding the Medico-Genomic Association”.

I would like to express my greatest gratitude to my supervisor, Prof. Dato’ Dr. Mohd Zaki Salleh as well as to my co-supervisor, Prof. Dr. Teh Lay Kek. Without their guidance, knowledge impartment and endless support throughout the whole course of my study, my knowledge on this subject would be nil. I would also like to thank them for their generosity in providing me the permission to use the facilities and reagents needed for this study which is available at the Integrative Pharmacogenomics Institute (iPROMISE). Thank you, Prof. Dato’ Dr. Mohd Zaki Salleh and Prof. Dr. Teh Lay Kek for pushing me to be better, keeping me on track and providing the necessary needs for this study. You have given me the opportunity to grow under your wings and for this I will never going to be able to repay you. You inspire me to further achieve more in life and to always dream big. Thank you, Prof!

I have particularly enjoyed working with the intelligent postgraduate students in iPROMISE in particular Lee Lian Shien, Mohd Shafiq bin Aazmi, and Muhamad Izwan Ismail. Thank you for all the insightful discussions. The advices, ideas, and constructive criticisms had surely helped me gain better insights of my study. I owe all of you so much and hope that someday I will be able to do the same for you. Also, I would like to thank all the members involved with the "Evolutionary Genomics and Anthropological Approaches on the Endangered Malaysian Aborigine Populations: Towards Ensuring their Sustainability" LRGS Grant who had helped out with the initial sample collection process. Without all your help, this study of mine would not even be possible.

Most of all, I would like to thank my mother and father, and Brigadier General Dato’ Ismet Nayan bin Ismail, for making the largest contribution of all. It was their unshakable faith in me that kept me going throughout this study.

Last but not least, I would also like to acknowledge the grant used for this study where the study was funded by the Ministry of Higher Education, Malaysia through the Long Term Research Grant (LRGS) scheme.

TABLE OF CONTENTS

CHAPTER ONE
INTRODUCTION

1.1 RESEARCH BACKGROUND

The Malaysian aborigines or also known as the "Orang Asli" are a collection of ethnic groups that are believed to be the indigenous peoples of Peninsular Malaysia. The Orang Asli comprises of three main ethno-linguistic tribes namely, the Senoi, Proto-Malays (or also known as the Aboriginal Malays) and the Negritos (Kasim, 2008; Khor & Shariff, 2008; Planning and Research Division JHEAO, 2008; Abu, Rahman, & Noor, 2013; Ahmad, Khalid, Quek, Anuar, & Phipps, 2013). Above that, each of these tribes consists of six different sub-tribes respectively. The Senoi is the largest tribe of the Orang Asli population, followed by the Proto-Malays, and lastly the Negritos (Kasim, 2008; Planning and Research Division JHEAO, 2008; Abu et al., 2013). Interestingly, data obtained from JHEOA in 2008 (Planning and Research Division JHEAO, 2008) shows that four of the Negritos sub-tribes: Kensiu, Kintak, Lanoh, and Mendriq; as well as the Orang Kanaq (from the Proto-Malay sub-tribe) and Che Wong (from the Senoi sub-tribe) are left with dwindling population numbers. This was also supported by a recent published article in 2013 (Abu et al., 2013) with similar pattern of population distribution. This leads to a question of what had caused these populations to remain small.

As up to date, there has not been any report on the whole genome sequences of the unique Orang Asli. Genetic related studies so far had focused more on their phylogeny (Macaulay et al., 2005; Hill et al., 2006; Peng et al., 2010; Ang et al., 2011; Jinam et al., 2012) and specific disease or trait related molecular conditions which include malaria-resistant genes (Bear, Lie-Injo, Welch, & Lewis, 1976), molecular basis of G6PD deficiency (Amini, Ismail, & Zilfalil, 2011; Amini & Ismail, 2013; Ismail et al., 2013), and skin colour distribution (Ang et al., 2012). This signifies the need for studies where whole genome sequencing is applied to generate the complete genomic architecture of the Orang Asli as unravelling the genomic architecture of the human DNA has been known to contribute to the understanding of the human evolution (Noonan, 2010; Jinam et al., 2012b; The 1000 Genomes Project Consortium, 2012;