UNIVERSITI PUTRA MALAYSIA

DEVELOPMENT OF MICROSATELLITE MARKERS AND GENETIC DIVERSITY ASSESSMENT OF KEMPAS (*KOOMPASSIA MALACCENSIS*) IN PENINSULAR MALAYSIA

LEE CHAI TING
IB 2009 19

DEVELOPMENT OF MICROSATELLITE MARKERS AND GENETIC DIVERSITY ASSESSMENT OF KEMPAS (*KOOMPASSIA MALACCENSIS*) IN PENINSULAR MALAYSIA

By

LEE CHAI TING

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Doctor of Philosophy

December 2009

Specially dedicated to my beloved husband and family members,
in loving memory of my late grandmother and uncle

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Doctor of Philosophy

DEVELOPMENT OF MICROSATELLITE MARKERS AND GENETIC DIVERSITY ASSESSMENT OF KEMPAS (*KOOMPASSIA MALACCENSIS*) IN PENINSULAR MALAYSIA

By

LEE CHAI TING

December 2009

Chairman: **Faridah Qamaruz Zaman, PhD**

Institute: **Bioscience**

A total of 24 novel microsatellite markers have been successfully isolated and characterised in an important tropical timber species of the family Leguminosae, *Koompassia malaccensis*, locally known as kempas. The microsatellite primers were designed from a genomic library enriched for dinucleotide (CT) repeats and subsequently screened on 24 samples from a natural population. In general, these microsatellite markers are highly polymorphic (mean number of alleles per locus, A_a = 6.84; average gene diversity, H_e = 0.692), with two loci found to deviate significantly from Hardy-Weinberg equilibrium ($p < 0.05$). The utility of these microsatellite markers were tested across 13 leguminous timber tree species and the highest transferability was found with *K. excelsa*, the only species of the same genus tested, followed by *Dialium platysepalum* of the same subtribe, Dialiinae. The amplification success appeared to be inversely associated with the phylogenetic distance, in particular up to the subtribal levels. Four of the microsatellite loci were used to study the mating system of *K. malaccensis*, based on a fruiting season at the

Semangkok Forest Reserve in year 2005. Single- and multilocus population outcrossing estimates (t_s and t_m) were determined using the program MLTR ver 3.0. The results showed that *K. malaccensis* is predominantly outcrossing (t_m = 0.890), with low tendency of mating between relatives [($t_m - t_s$) = 0.027]. In addition, the level of genetic diversity of *K. malaccensis* in 34 natural populations throughout Peninsular Malaysia was assessed and its distribution described. Omitting four loci due to suspected presence of null alleles and linkage disequilibrium, 20 microsatellite loci were analysed for 974 individuals. Overall, all the populations showed high levels of genetic diversity, with gene diversity (H_e) ranging from 0.577 (Kuala Langat Selatan) to 0.787 (LenggorB) and mean A_a being 9.0. The levels of genetic diversity for the two peat swamp (PS) populations (Kuala Langat Selatan and Pekan) were significantly lower than for the non-PS populations. The estimated coefficients of population differentiation (F_{ST} and R_{ST}) revealed that the majority of the genetic diversity resides within populations and less among populations (F_{ST}: 0.077; R_{ST}: 0.102). Results from the analysis of molecular variance (AMOVA), cluster analysis, principal component analysis (PCA) and STRUCTURE analysis consistently demonstrated that *K. malaccensis* originating from the two contrasting habitats (PS vs non-PS) were genetically distinct, supporting the ecotype hypothesis. Excluding the PS populations, the among-population component of genetic diversity was even smaller (F_{ST}: 0.028; R_{ST}: 0.023), but statistically significant. Pairwise F_{ST} values among the non-PS populations were positively correlated to geographical distance (Mantel test; r^2 = 0.0936, $p < 0.01$), indicating weak but significant isolation-by-distance. Pangkor Selatan and LenggorB were found to be relatively more divergent among the non-PS populations investigated, presumably due to genetic drift and the inclusion of freshwater swamp habitat, respectively. Significant but weaker

population genetic structure was detected among the rest of the non-PS populations surveyed, which corresponded to the topography of Peninsular Malaysia, reflecting the role of mountain ranges as geographical barriers to gene flow. The implications of the findings from this study for the genetic conservation of *K. malaccensis* are discussed and conservation strategies (both *in situ* and *ex situ*) proposed to ensure sustainable utilisation of this important timber species in Malaysia.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan ijazah Doktor Falsafah

PEMBANGUNAN PENANDA MIKROSATELIT DNA DAN PENILAIAN KEPELBAGAIAN GENETIK BAGI KEMPAS (*KOOMPASSIA MALACCENSIS*) DI SEMENANJUNG MALAYSIA

Oleh

LEE CHAI TING

Disember 2009

Pengerusi: **Faridah Qamaruz Zaman, PhD**

Institut: **Biosains**

Sejumlah 24 penanda mikrosatelit baru telah berjaya dipencil serta dicirikan untuk satu spesies balak tropika yang penting daripada famili Leguminosae, iaitu *Koompassia malaccensis*, yang dikenali dengan nama tempatan kempas. Pencetus mikrosatelit tersebut telah direka daripada perpustakaan genomik yang diperkayakan dengan ulangan dinukleotida (CT) dan kemudiannya disaring menggunakan 24 sampel dari satu populasi semulajadi. Secara amnya, penanda-penanda mikrosatelit yang diperolehi adalah berpolimorfik tinggi (min bilangan alel setiap lokus, A_a = 6.84; min kepelbagaian gen, H_e = 0.692), dengan dua lokus didapati menyimpang daripada keseimbangan Hardy-Weinberg ($p < 0.05$). Kegunaan penanda-penanda molekul tersebut diuji ke atas 13 spesies balak legum dan kadar pemindahan tertinggi didapati pada *K. excelsa*, satu-satunya spesies daripada genus yang sama yang telah diuji, diikuti dengan *Dialium platysepalum* daripada subtrib yang sama, Dialiinae. Kejayaan amplifikasi didapati mempunyai perhubungan songsang dengan jarak filogenetik, terutamanya sehingga ke peringkat subtrib. Empat daripada lokus

mikrosatelit tersebut telah digunakan untuk kajian sistem kacukan *K. malaccensis*, berdasarkan suatu musim buah di Hutan Simpan Semangkok pada tahun 2005. Anggaran kadar kacukan luar populasi berdasarkan lokus tunggal serta gabungan lokus (t_s and t_m) telah ditentukan dengan menggunakan program MLTR ver 3.0. Keputusan menunjukkan bahawa *K. malaccensis* mengamalkan kacukan luar dengan kadar yang tinggi (t_m = 0.890) dan mempunyai kadar kacukan sesama saudara yang rendah [($t_m - t_s$) = 0.027]. Di samping itu, tahap kepelbagaian genetik *K. malaccensis* bagi 34 populasi semulajadi di Semenanjung Malaysia telah dinilai dan taburannya diterangkan. Setelah mengasingkan empat lokus yang disyaki mempunyai alel-nul serta ketakseimbangan rangkaian, 20 lokus mikrosatelit telah dianalisis untuk 974 individu. Secara keseluruhannya, kesemua populasi menunjukkan tahap kepelbagaian genetik yang tinggi, dengan kepelbagaian gen (H_e) menjulat daripada 0.577 (Kuala Langat Selatan) ke 0.787 (LenggorB) dan min A_a sebanyak 9.0. Tahap kepelbagaian genetik untuk kedua-dua populasi paya gambut (Kuala Langat Selatan dan Pekan) adalah lebih rendah secara statistik berbanding dengan populasi bukan-paya-gambut yang lain. Anggaran koefisien pembezaan populasi (F_{ST} dan R_{ST}) menunjukkan bahawa majoriti kepelbagaian genetik dibahagikan di dalam populasi dan kurang di kalangan populasi (F_{ST}: 0.077; R_{ST}: 0.102). Keputusan daripada analisis varians molekular (AMOVA), analisis kelompok, analisis komponen prinsipal (PCA) serta analisis STRUCTURE secara konsisten menunjukkan bahawa *K. malaccensis* yang berasal daripada dua habitat yang berlainan (paya gambut dan bukan-paya-gambut) adalah berbeza secara genetik, iaitu menyokong hipotesis ekotip. Dengan pengecualian populasi paya gambut, komponen kepelbagaian genetik di kalangan populasi adalah jauh lebih kecil (F_{ST}: 0.028; R_{ST}: 0.023), tetapi signifikan secara statistik. Nilai F_{ST} secara berpasangan di kalangan populasi bukan-

paya-gambut berhubungan secara langsung dengan jarak geografi masing-masing (Mantel test; $r^2 = 0.0936$, $p < 0.01$), menggambarkan wujudnya pengasingan-oleh-jarak yang lemah tetapi signifikan. Pangkor Selatan dan LenggorB didapati lebih mencapah secara relatif di antara populasi bukan-paya-gambut yang dikaji, kemungkinan disebabkan akibat hanyutan genetik serta perangkuman habitat paya-air-tawar. Struktur genetik populasi yang lebih lemah tetapi signifikan telah dikesan di kalangan populasi bukan-paya-gambut yang lain, dan didapati berkait rapat dengan topografi Semenanjung Malaysia, menggambarkan peranan banjaran pergunungan sebagai halangan geografi terhadap aliran gen. Implikasi penemuan kajian ini terhadap pemuliharaan genetik telah dibincang dan strategi pemuliharaan (secara *in situ* dan *ex situ*) disyorkan demi memastikan penggunaan mampan spesies balak yang penting ini di Malaysia.

ACKNOWLEDGEMENTS

Praise be to God, Maker of all things. I give thanks to the Almighty God for His faithfulness and amazing grace in seeing me through. It has been a challenging but enriching and memorable journey.

I would like to express my heartfelt gratitude and appreciation to my supportive and understanding supervisors, Associate Professor Dr. Faridah Qamaruz Zaman, Dr. Lee Soon Leong and Professor Dr. Siti Shapor Siraj, for their valuable guidance, insightful ideas, constructive suggestions, great patience and constant encouragement throughout the course of this study.

A note of thanks to Dr. Saw Leng Guan, Dr. William Goodwin, Dr. Iyengar Arati, Dr. Naoki Tani, Dr. Yoshihiko Tsumura, Dr. Saneyoshi Ueno and Dr. Remy Petit for sharing their expertise and sparking ideas through constructive discussions. As ultraspeed centrifuge was not available in FRIM, the CsCl purification of the genomic DNA for microsatellite development was performed with the help from Dr. Tani at the Forestry and Forest Products Research Institute (FFPRI), Tsukuba, Japan. Many thanks go to Mr. Mohd Nor Mat Isa from the Malaysia Genome Institute for his assistance in sequence analysis and to Mr. Chua Bok Hui for introducing the MISA software. I would also like to extend my sincere thanks to my lab-mates, Dr. Kevin Ng, for lending a big hand during field samplings and for sharing his technical skills generously, and Mrs. Tnah Lee Hong for helping me to familiarise with some of the software used.

I am indebted to FRIM's Director General, Dato' Dr. Abdul Latif Mohmod, and the ex-Director General, Dato' Dr. Abdul Razak Mohd Ali, for giving me the opportunity for further study. I am also thankful to the Forest Biotechnology Division Director, Dr. Norwati Muhammad, and the previous Senior Directors, Dr. Marzalina Mansor and Dr. Daniel Baskaran Krishnapillay, for they have been very supportive and understanding. My sincere appreciation is extended to my senior colleagues, Mr. Ang Khoon Cheng, Mr. Lim Seng Choon, Dr. Lillian Chua, Dr. Abdul Rahman Kassim and Dr. Ismariah Ahmad for being helpful and approachable. This research is funded by the Ministry of Science, Technology and Innovation Malaysia (IRPA Project No. 09-04-01-0098-EA001). The scholarship from the Public Services Department (JPA), the financial support from FRIM, as well as a token of graduate research grant from the ASEAN-Korea Environmental Cooperation Project (AKECOP) are gratefully acknowledged.

All field samplings were carried out with the help of the dedicated research assistants from the Genetic Laboratory, FRIM (Mr. Ramly Punyoh, Mr. Ghazali Jaafar, Mr. Yahya Marhani and Mrs. Sharifah Talib), besides Dr. Lee Soon Leong and Dr. Kevin Ng. Without their help, this study would not have been possible. It has been a challenging but rewarding experience, collecting samples from various forest reserves. I would also like to thank the Forest Department Peninsular Malaysia and the Forest Departments of the respective states for granting permission for sample collection. The kind logistic assistance of the District Forest Officers, Assistant District Forest Officers, foresters, and rangers is much appreciated. Special acknowledgement is also due to my laboratory's research assistants for their help in DNA extraction (Mrs. Mariam Din, Miss Suryani Che Semat, Mrs. Nurul Hudaini

Mamat, Mrs. Nor Salwah Abd Wahid and those mentioned above). To all the other colleagues from the Genetic Laboratory (Dr. Norwati Adnan, Dr. Norlia Basherudin, Dr. Ng Chin Hong, Dr Siti Salwana Hashim and Dr. Mohd Rosli Haron), thanks for maintaining the conducive working environment.

I am deeply grateful to my mom and dad for encouraging me to further my studies. I am thankful for all my family members and friends (Wai Mun, Phoon, How, Sik, Lee Hong, Wan Fong, Brian, Tzer Ying, etc) who cheer me up with their moral support and kept me in prayers. Finally, I want to thank my dear husband, Yik Sheng, for his love, support, understanding and encouragement. I will always cherish the joys and tears we shared in pursuing our Ph.D. at the same. His grace is sufficient for us. Hallelujah!

This thesis submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Faridah Qamaruz Zaman, PhD
Associate Professor
Institute of Bioscience
Universiti Putra Malaysia
(Chairman)

Siti Shapor Siraj, PhD
Professor
Faculty of Agriculture
Universiti Putra Malaysia
(Member)

Lee Soon Leong, PhD
Forest Biotechnology Division
Forest Research Institute Malaysia
(Member)

HASANAH MOHD GHAZALI, PhD.
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 17 Mac 2010

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Koompassia malaccensis (Northwest, NS; Southwest, SW; North & Central, NC; East & South, ES) as identified from the cluster analysis (see Figure 4.10).

LIST OF ABBREVIATIONS

AMOVA	Analysis of molecular variance
bp	Base pair
CAPS	Cleaved amplified polymorphic sequences
AFLP	Amplified fragment length polymorphism
CITES	Convention on International Trade in Endangered Species
CpDNA	Chloroplast DNA
CsCl	Caesium chloride
CTAB	Hexadecyltrimethyl-ammonium bromide
DAF	DNA amplification fingerprint
dbh	Diameter at breast height
DNA	Deoxyribonucleic acid
dNTP	2'-deoxynucleoside 5'-triphosphate
EDTA	Diaminoethanetetra-acetic acid
EMBL	European Molecular Biology Laboratory
EST	Expressed sequence tag
ESUs	Evolutionarily significant units
FAO	Food and Agriculture Organization of the United Nations
FIASCO	Fast isolation by AFLP sequences containing repeats
FR	Forest Reserve
FRIM	Forest Research Institute Malaysia
GSM	Generalized stepwise model
IAM	Infinite alleles model
IBD	Identical-by-descent