ISOLATION AND CHARACTERIZATION OF MICROSATELLITE LOCI IN AN ENDANGERED PALM, *JOHANNESTEIJSMANNIA LANCEOLATA* (ARECACEAE)

ANG CHENG CHOON

FACULTY OF SCIENCE

UNIVERSITY OF MALAYA

KUALA LUMPUR

2012

ISOLATION AND CHARACTERIZATION OF MICROSATELLITE LOCI IN AN ENDANGERED PALM, *JOHANNESTEIJSMANNIA LANCEOLATA* (ARECACEAE)

ANG CHENG CHOON

DISSERTATION SUBMITTED IN FULFILMENT OF THE REQUIREMENT FOR THE DEGREE OF MASTER OF SCIENCE

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2012

ABSTRACT

Johannesteijsmannia lanceolata J. Dransf. (Arecaceae) is a rare and endangered palm species which is endemic to Peninsular Malaysia. It was listed as an endangered species in the 1997 IUCN red list of threatened plants owing to its substantially reduced population size. Hence, understanding of the population structure and genetic diversity by using microsatellite markers is important to postulate scientific conservation and restoration strategies for this species.

In this study, a set of 33 novel polymorphic microsatellite loci had been successfully isolated and characterized for *J. lanceolata* from a genomic library enriched for AG/CT repeats via the selective hybridization of biotinylated $(CT)_{15}$ probes and streptavidin-coated magnetic beads. The microsatellite primers were screened on 24 samples collected fresh from the Angsi Reserve Forest. In general, the microsatellite loci exhibited high level of polymorphism with an average number of 10 alleles per locus and mean polymorphic information content (PIC) of 0.771. Two loci (*Jla*130b and *Jla*332a) showed significant departures from the Hardy-Weinberg equilibrium with ($P < 1.515 \times 10^{-3}$) after Bonferroni correction. Micro-checker analysis suggested the presence of null alleles in both the *Jla*332a and *Jla*118 loci. No significant linkage disequilibrium was detected between any pair of loci. All 33 microsatellite loci were confirmed to be conforming to the Mendelian mode of inheritance in a test with an open-pollinated half-sib family where all the progenies inherited at least one maternal allele. Transferability of the microsatellite loci were checked across other *Johannestsijsmannia* species. All loci showed positive amplifications in four samples from each of the three *Johannesteijsmannia* species, except for loci *Jla*124 (*J. magnifica* J. Dransf.) and *Jla*168b (*J. magnifica* and *J. perakensis* J. Dransf.).

Of the 33 newly isolated microsatellite loci, twenty-eight loci are in Hardy-Weinberg equilibrium, highly polymorphic, absence of linkage disequilibrium and null alleles. Therefore, they could be used as markers for the investigation of the population genetic structure, gene flow, parentage, and mating system of *J. lanceolata* across its distribution range in the near future.

ABSTRAK

Johannesteijsmannia lanceolata J. Dransf. (Arecaceae) merupakan spesis palma yang jarang ditemui dan endemik di Semenanjung Malaysia. Ia telah didaftarkan sebagai tumbuhan terancam oleh '*1997 IUCN red list of threatened plants*' berdasarkan saiz populasinya yang semakin berkurangan. Oleh itu, pemahaman tentang struktur populasi dan diversiti genetik dengan menggunakan penanda mikrosatelit amat diperlukan untuk proses konservasi dan sebarang strategi bagi pemulihan spesies tersebut.

Dalam penyelidikan ini, sebanyak 33 lokus microsatelit baru yang polimorfik telah berjaya diasingkan dan dicirikan untuk *J. lanceolata* daripada sumber genomik yang diperkaya dengan ulangan AG/CT melalui hibridisasi penunjuk biotinylated $(CT)_{15}$ dengan manik-manik magnet yang dilapisi streptavidin. Primer mikrosatelit telah disaring daripada 24 sampel dari Hutan Simpan Angsi, Negeri Sembilan. Secara umumnya, lokus mikrosatelit tersebut menunjukkan tahap polimorfik yang tinggi dengan bilangan purata 10 alel bagi setiap lokus dan purata 0.771 kandungan informasi polimorfik (PIC). Dua lokus (*Jla*130b dan *Jla*332a) telah menunjukkan keberangkatan yang signifikan dari Keseimbangan Hardy-Weinberg di mana $P < 1.515 \times 10^{-3}$ setelah pembetulan Bonferroni. Analisis perisian Micro-checker mencadangkan kehadiran alel null dalam kedua-dua lokus *Jla*118 dan *Jla*332a pada aras kepercayaan 99 %. Tiada sebarang hubungan yang tidak seimbang yang signifikan didapati antara setiap pasangan lokus tersebut. Kesemua 33 lokus mikrosatelit ini mematuhi mod warisan Mendelian melalui ujian dengan keluarga separa-beradik-pendebungaan terbuka di mana semua progeni mewarisi sekurang-kurangnya satu alel ibu. Lokus amplifikasi bersilang mikrosatelit diperiksa bagi semua spesis dalam genus *Johannestsijsmannia*. Semua lokus telah menunjukkan amplifikasi positif dalam ketiga-tiga

spesies *Johannesteijsmannia* yang lain kecuali lokus *Jla*124 (*J. magnifica* J. Dransf.) dan lokus *Jla*168b (*J. magnifica* dan *J. perakensis* J. Dransf.).

Daripada 33 penanda mikrosatelit baru ini, 28 lokus berada dalam keseimbangan Hardy-Weinberg, sangat polimorfik, ketiadaan ketakseimbangan hubungan dan alel null. Oleh itu, mereka boleh digunakan sebagai penanda untuk mengkaji taburan populasi spesis ini dari segi struktur genetik populasi, aliran gen, keturunan, dan sistem kacukan *J. lanceolata* pada masa akan datang.

ACKNOWLEDGEMENTS

I would like to thank my supervisors, Dr. Ng Ching Ching and Associate Professor Dr. Rozainah Mohamad Zakaria for their advice, tremendous support, constant encouragement, and valuable criticism during the preparation of this thesis.

I wish to express my gratitude and appreciation to Genetics Laboratory of Forest Research Institute Malaysia (FRIM) for providing the place and facilities to conduct this project. Many thanks go to Dr Lee Soon Leong for his valuable advice, constructive suggestions and critical comments on the preparation of the manuscript during paper submission. I would like to extend my appreciation to Dr Lee Chai Ting and Dr Tnah Lee Hong for their guidance, patient and assistance during the construction of enriched-library. Not forgetting to thank Dr. Kevin Ng Kit Siong for sharing his knowledge on fragment analysis, Dr. Ng Chin Hong for introducing the MISA software, Ms Chan Yoke Mui for describing the ecology and morphological features of *Johannesteijsmannia lanceolata*, Dr. Naoki Tani for sharing his valuable experience and some papers by him and also members of the Genetics Laboratory for being unbelievably helpful during the total DNA extraction.

I greatly acknowledge the Ministry of Agriculture and Agro-Based Industry, Malaysia (Science Fund No: 05-01-03-SF0199 & 05-03-10-SF1031) and postgraduate research fund (PS228/2009C) for funding this project. Besides, I also appreciate the University of Malaya Fellowship Scheme for its financial support.

Special thanks to my fellow lab mates for their friendship, encouragement, support and care. My gratitude also goes to my friends whom I share my joys and grieves with. Last but not least, I want to dedicate this thesis to my family for their lifetime of love and unconditional support.

PUBLICATION AND PRESENTATION

Paper publication:

Ang, C. C., Lee, S. L., Lee, C. T., Tnah, L. H., Zakaria, R. M., & Ng, C. C. (2011). Isolation and characterization of microsatellite loci in an endangered palm, *Johannesteijsmannia lanceolata* (Arecaceae). *American Journal of Botany, 98*(5), e117-e119.

Poster presentation:

Ang Cheng Choon, Ng Ching Ching, Lee Soon Leong, Lee Chai Ting, Tnah Lee Hong and Rozainah Mohamad Zakaria, 2010. **Isolation and characterization of microsatellite loci of *Johannesteijsmannia lanceolata* (Arecaceae)**. My1bio conference 2010: Accelerating biotechnology through innovative research, 30-31 Oct 2010, Berjaya Times Square Hotel, Kuala Lumpur, Malaysia. (National)

TABLE OF CONTENTS

ABSTRACT i

ABSTRAK iii

ACKNOWLEDGEMENTS v

PUBLICATION AND PRESENTATION vi

TABLE OF CONTENTS vii

LIST OF FIGURES x

LIST OF TABLES xi

LIST OF SYMBOLS AND ABBREVIATIONS xii

CHAPTER

1 INTRODUCTION 1

2 LITERATURE REVIEW 4

2.1 The genus, *Johannesteijsmmania* 4

2.1.1 Biology and Ecology 6

2.1.2 Distribution 6

2.2 *Johannesteijsmannia lanceolata* *7*

2.3 Genetic markers 9

2.3.1 DNA-based molecular markers 9

2.3.1.1 Restriction Fragment Length Polymorphism, RFLP 11

2.3.1.2 Random Amplified polymorphic DNA, RAPD 11

2.3.1.3 Amplified Fragment Length Polymorphism, AFLP 12

2.3.1.4 Single Nucleotide Polymorphism, SNP 12

2.3.1.5 Expressed Sequences Tagged, EST 13

2.3.1.6 Microsatellites 13

2.3.1.6.1 Classification of microsatellites 14

2.3.1.6.2 Mutation mechanisms of microsatellites 14

2.4 Applications of microsatellite markers to plant population analysis 16

2.4.1 Population structure assessment 16

2.4.2 Parentage and paternity analysis 17

2.4.3 Gene flow 17

2.5 Development of microsatellites 18

2.6 Cross-species amplification 19

3 MATERIALS AND METHODS 21

3.1 Leaf samples collection 21

3.2 Total genomic DNA extraction 21

3.3 Construction of CT-enriched library 22

3.3.1 Digestion of genomic DNA with restriction enzyme 22

3.3.2 Ligation of DNA fragments with cassettes 23

3.3.3 PCR amplification to confirm ligation 24

3.3.4 Selective hybridization using biotinylated $(CT)_{15}$ 25

3.3.5 Digestion of cassettes 27

3.3.6 Ligation of CT-enriched DNA into plasmid 27

3.3.7 Transformation of recombinant plasmids into competent cells 28

3.4 Amplification of circular DNA 28

3.4.1 DNA sequencing and analysis of sequence data 28

3.5 Primer design 29

3.6 Primer screening and PCR amplification 30

3.7 Fragment analysis 30

3.8 Characterization of microsatellite loci 30

3.9 Analysis of inheritance 31

3.10 Cross-species amplification 32

4 RESULTS 33

4.1 Construction of CT-enriched library 33

4.2 DNA sequencing of positive clones 38

4.3 Primer design 42

4.4 Primer screening 42

4.5 Characterization of microsatellite loci 44

4.6 Analysis of inheritance 51

4.7 Cross-species amplification 53

5 DISCUSSION 54

5.1 Construction of enriched library 54

5.2 Analysis of sequence data 57

5.3 Primer design 58

5.4 Primer screening 58

5.5 Characterization of microsatellite loci 59

5.6 Cross-species amplification 64

5.7 Future plans 65

6 CONCLUSION 66

REFERENCES 67

APPENDICES 76

APPENDIX A: DNA sequences derived from microsatellite-containing clones of *Johannesteijsmannia lanceolata* with respective clone name and followed by the NCBI GenBank Accession number 76

APPENDIX B: Paper published in American Journal of Botany Primer Notes & Protocol in the Plant Sciences 82

LIST OF FIGURES

Figure		Page
2.1	*Johannesteijsmannia lanceolata*	7
2.2	Model of microsatellites mutation by DNA replication slippage adapted from Oliveira et al. (2006)	15
3.1	A: Longitudinal section of fruit of *J. lanceolata* with embryo in the red circle. B: Embryo	31
4.1	Ethidium bromide stained 0.85 % (w/v) agarose gel of extracted genomic DNA	33
4.2	Ethidium bromide stained 1.6 % (w/v) agarose gel of digested genomic DNA with *Nde* II	34
4.3	Ethidium bromide stained 1.6 % (w/v) agarose gel of PCR amplicons	35
4.4	Ethidium bromide stained 1.6 % (w/v) agarose gel of enriched DNA fragments after removal of cassettes	36
4.5	A LB agar plate containing ampicillin, IPTG and X-gal showing the growth of transformed white bacterial colonies after overnight incubation at 37 ºC	37
4.6	Chromatograms showing microsatellite sequences from clone A, *Jla*048: $(GA)_{15}$ repeat; clone B, *Jla*187a: $(TG)_8$ repeat and clone C, *Jla*332a: $(CT)_{14}$ repeat	39
4.7	Ethidium bromide stained 2.0 % (w/v) agarose gel of PCR products generated using four primer sets	43
4.8	Electropherograms of polymorphic alleles amplified from four individuals of *J. lanceolata* using primer pairs A, *Jla*117; B, *Jla*112 and C, *Jla*160	44
4.9	Electropherograms of allele's segregation showed in a half-sib family at locus *Jla*168b	52
4.10	Ethidium bromide stained 2.0 % (w/v) agarose gel of specific PCR amplicons generated from the positive cross-species amplification of locus *Jla*106 in three *Johannesteijsmannia* species	53
5.1	Relationships between length of the microsatellite repeat motifs and number of alleles observed per locus	62
5.2	Relationships between length of microsatellite repeat motifs and PIC values	63

LIST OF TABLES

Table		Page
2.1	Overview of the relevant characteristics of molecular markers adapted from Ng (2005)	10
3.1	Preparation for DNA digestion with *Nde* II	22
3.2	Ligation set up of DNA fragments with *Sau*3A1 cassettes	23
3.3	Repair of nicks between DNA fragments and cassettes	24
3.4	PCR reaction mixture	25
3.5	Hybridization of cassette-ligated genomic DNA and biotinylated $(CT)_{15}$	26
3.6	Digestion of cassettes mixture	27
3.7	Ligation of enriched DNA into plasmid	28
4.1	Total numbers of microsatellites identified according to the respective unit sizes	40
4.2	Frequency of different microsatellites repeat motifs identified from the CT-enriched genomic library of *J. lanceolata*	41
4.3	Characteristics of the 33 polymorphic microsatellite loci which include locus name, repeat motifs, forward (F) and reverse (R) primer sequences, annealing temperatures (T_a), number of alleles per locus (A), allele size, observed (H_O) and expected (H_E) heterozygosities and polymorphic information content (PIC) and GenBank accession numbers	46

LIST OF SYMBOLS AND ABREVIATIONS

%	Percentage
′	Minute
″	Second
°	Degree
µg	Microgram
µL	Microliter
µm	Micrometer
µM	Micromolar
2n	Diploid
A	Adenine
AFLP	Amplified fragment length polymorphism
bp	Base pair
BSA	Bovine serum albumin
C	Cytosine
cDNA	Complementary DNA
cm	Centimeter
CTAB	Cetyl trimethyl ammonium bromide
dH_2O	Distilled water
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
E	East
EDTA	Ethylenediaminetetraacetic acid
EST	Expressed sequences tagged
FIASCO	Fast isolation by AFLP sequences containing repeats
G	Guanine
g	Gram

H_E	Expected heterozygosity
H_O	Observed heterozygosity
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IUCN	International union for conservation of nature
kcal/mol	Kilocalorie per mole
KCl	Potassium chloride
LB	Luria bertani
m	Proportion of migrants per generation
M	Molar
mg	Milligram
$MgCl_2$	Magnesium chloride
min	Minute
MISA	Microsatellite identification tool
mL	Milliliter
mM	Milimolar
N	North
N	Population size
NaCl	Sodium chloride
NCBI	National Center for Biotechnology Information
ng	Nanogram
Nm	Gene flow
nm	Nanometer
ºC	Degree Celcius
OD	Optical density
PCR	Polymerase chain reaction
pg	Picogram
PIC	Polymorphic information content
PIMA	PCR isolation of microsatellites array

pmol	Picomole
PVP-40	Polyvinylpyrrolidone average mol wt 40,000
r	Pearson's correlation coefficient
RAPD	Random amplified polymorphic DNA
RE	Restriction enzyme
RFLP	Restriction fragment length polymorphism
rpm	Revolutions per minute
SDS	Sodium lauryl sulfate
sec	Second
SNP	Single nucleotide polymorphism
SSC	Saline-sodium citrate
SSLP	Simple sequence length polymorphism
SSR	Simple sequence repeat
STR	Short tandem repeat
T	Thymine
T_a	Annealing temperature
TAE	Tris acetate EDTA
TE	Tris EDTA
T_m	Melting temperature
Tris-Cl	Tris(hydroxymethyl)aminomethane adjust pH with hydrochloric acid
U	Unit
UTR	Untranslated region
VNTR	Variable number of tandem repeat
w/v	Weight per volume
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
β	Slope coefficient

CHAPTER 1
INTRODUCTION

CHAPTER 1

INTRODUCTION

The palm family, Arecaceae or Palmae includes about 2,364 species in 190 genera which are generally distributed throughout the tropics and subtropics of America and Asia (Govaerts and Dransfield, 2005). Malaysia is one of the blessed countries in the world with extremely high diversity of palm species with 398 species of indigenous palms in 33 genera (Saw, 1998). Nevertheless, some of the palms are highly endangered due to human activities. Of the 38 highly endangered endemic species in Malaysia, 21 of them are found only in Peninsular Malaysia (Johnson, 1996) and *Johannesteijsmannia lanceolata* J. Dransf. is one of the highly endangered palms species listed in the 1997 IUCN red list of threatened plants (Walter and Gillett, 1998). *J. lanceolata* is a stunning stemless or short-stemmed under storey palm in the primary forest in Peninsular Malaysia. It can be easily recognized by its long tapering undivided leaf and the corky warted fruits (Dransfield et al., 2008). However, populations of this species are declining due to loss of habitat because of deforestation for development and agriculture, logging, over-collection of leaves and seeds notwithstanding that the species is indeed susceptible to any forest disturbance.

Several phenological studies on the flowering event, vegetative growth, population structure and spatial distribution of this species have been done to investigate the plant population (Rozainah and Sinniah, 2005; Rozainah and Sinniah, 2006). In 2009, Chan studied the reproductive biology and ecology of *J. lanceolata* likewise. Both studies revealed the critical situation of less (2-6) mature fruits production per fruiting season in a long reproductive cycle. To date, there is insufficient genetic data published regarding this species. Therefore, this study aimed to develop microsatellite markers for future population genetic studies of *J. lanceolata*.

Microsatellite is also known as simple sequence repeat (SSR) which consists of multiple copies of tandemly repeated motif of 1-6 bp widely intersperse throughout the genomes of living organisms (Litt and Luty, 1989; Tautz, 1989). Microsatellites are widely used as genetic marker due to the features of codominant, multiallelic, highly polymorphic, abundant, reproducible, selective neutrality, and easily accessible to PCR amplification (Fisher et al., 1996). Furthermore, cross-species transferability of microsatellite markers also made them an ideal tool for most plant genetic studies (Peakall et al., 1998). Hence, it is feasible to use microsatellite as molecular marker for the future population structure and genetic diversity assessment of *J. lanceolata* towards long-term goal of postulating scientific conservation and restoration strategies for the species.

OBJECTIVES

The development of microsatellite markers is crucial for population studies of endangered species. Genetic diversity and population structure of the species can be assessed with microsatellite markers. Therefore, this study aimed to isolate and characterize polymorphic microsatellite loci as the molecular markers to be used in future population genetic studies. Ultimately, more effective conservation strategies can be postulated to assist forest management in the genetic resource conservation of this endangered palm species. The specific objectives of this study were:

1. To isolate polymorphic microsatellite markers from CT-enriched library of *J. lanceolata*.
2. To characterize the polymorphic microsatellite loci by screening with samples of *J. lanceolata* from the Angsi Reserve Forest.
3. To check cross-species amplifications of the polymorphic microsatellite loci in three other species of *Johannesteijsmannia*.

CHAPTER 2
LITERATURE REVIEW

CHAPTER 2

LITERATURE REVIEW

2.1 The genus, *Johannesteijsmmania*

Johannesteijsmannia belongs to the order Principes, family Arecaceae or Palmae, subfamily Coryphoideae, tribe Corypheae and subtribe Livistoninae (Uhl and Dransfield, 1987). The genus was named after Johannes Elias Teijsmann (1808–1882), a Dutch gardener and botanist at the Buitenzorg Botanical Garden, Java (now Kebun Raya Indonesia, Bogor). Local people in Peninsular Malaysia called them as Daun Payung, Sal, Sang, Koh (Dransfield, 1970) and Umbrella Leaf Palm as a single leaf of *Johannesteijsmmania* is capable to be an excellent umbrella. The leaves have been used for thatch and shelters among the aboriginal communities (orang asli) in Peninsular Malaysia. In Kelantan, fruits of *J. altifrons* (Reichb. f. et Zoll.) H. E. Moore collected from the wild was used in Chinese herbal medicine (Kiew, 1991).

In 1972, Dransfield did a comprehensive investigation on the species of *Johannesteijsmmania*. He recognized three new species of *Johannesteijsmmania* (other than *J. altifrons*) in Malaysia: *J. perakensis* J. Dransf., *J. magnifica* J. Dransf., *J. lanceolata*. He described the genus as moderate, solitary, armed, acaulescent to short-trunked, pleonanthic or polycarpic, hermaphroditic palm. Below are the characteristics of *Johannesteijsmmania*:

- **Stem** is short, decumbent, ringed with close leaf scars.
- **Leaves** are large, entire, undivided, diamond-shaped, subpinnately ribbed, can reach up to 6-7 meters in length, marcescent, leaf sheath formed tubular and separated into an interwoven mass of fibres as they grow older; petiole well developed, relatively triangular in cross-section, adaxially flattened, armed along the margins with small, sharp teeth, leaf blade subpinnately ribbed, glabrous or the abaxial

surface covered with white indumentums especially in *J. magnifica*, and irregularly stepped appearance in the margin.

- **Inflorescences** are interfoliar, short, partly buried by leaf litter, 1–5 branching orders, up to 7 in number, colour changed from cream to cinnamon-brown, tubular, relatively inflated.
- **Flowers** are cream-coloured, strongly scented, sessile, calyx consisted of 3 petals and 3 sepals stamens 6, epipetalous, filaments very broad, fleshy, angled, unexpected short, slender, distinct tips, small anthers, rounded.
- **Pollen** is bi-symmetric or slightly asymmetric, scabrate, perforate, longest axis 20–32 μm.
- **Fruits** are chestnut brown in colour, rounded, developing from 1 carpel rarely 2 or 3 carpels, the fruit then 2 or 3-lobed; epidermis died in the early of development leaving mesocarp to crack and generate thick, corky, pyramidal warts at maturity; endocarp moderately thick, crustaceous.
- **Seed** is basally attached, endosperm homogeneous, embryo lateral. Germination remote-tubular.
- **Cytology**: 2n = 34.

2.1.1 Biology and Ecology

These extraordinary palms are frequently abundant and dominate the undergrowth of primary rain forest. They are never found in secondary forest owing to their intolerant to any disturbance. They are also hardly found in wet valley bottom or swamp soils. *J. magnifica* and *J. lanceolata* are commonly found in hill slopes whereas *J. perakensis* is frequently grows at hill slopes and ridge tops. Distribution of these palms is surprisingly disjunctive and hard to know the reason of such pattern (Dransfield et al., 2008).

2.1.2 Distribution

J. altifrons is widely spread in south Thailand, West Malaysia, Sumatra, and the western part of Borneo whereas the other three species are endemic to Peninsular Malaysia. The magnificent palm, *J. magnifica* has been recorded growing in Sungai Lalang Forest Reserve, Semenyih (3° 03′ 26.7″N, 101° 51′ 16.2″E) in Selangor and Berembun Forest Reserve, Bukit Tangga (2° 51′ 37.7″N, 102° 01′ 00.0″E) in Negeri Sembilan (Dransfield, 1972). The narrow leaves palm, *J. lanceolata* has been recorded in Sungai Lalang Forest Reserve, Semenyih (3° 03′ 29.1″N, 101° 52′ 22.4″E) in Selangor, growing together with *J. magnifica*, Ulu Sungai Tekal Besar, Temerloh (3° 43′ 01.0″N, 102° 16′ 52.0″E) in Pahang and Gunung Angsi Forest Reserve (1° 57′ 39.5″N, 99° 18′ 06.4″E) in Negeri Sembilan. The trunked palm, *J. perakensis* has been recorded in Kledang Saiong Forest Reserve (4° 42′ 57.24″N, 100° 58′ 2.16″E) and Gunung Bubu Forest Reserve (4° 33′ 38.6″N, 100° 51′ 07.7″E) in Perak.

2.2 *Johannesteijsmannia lanceolata*

J. lanceolata is easy to be recognized from the other three species by its distinctive features which are the long tapering slender leaves (Figure 2.1 A & B), infructescences with 3-6 thick branches bearing spirals of papillate-petalled flowers (Figure 2.1 C) and corky warted fruits (Figure 2.1 D & E). *J. magnifica* and *J. altifrons* display a much larger number of inflorescence branches with 1000 and 100 respectively whereas *J. perakensis* which is very similar to *J. altifrons* but differs in its ascending stem that formed a trunk up to 4 meters tall. Another feature that distinguished *J. magnifica* from the other three species is the dense grey-white indumentums covered on the below surface of the leaf.

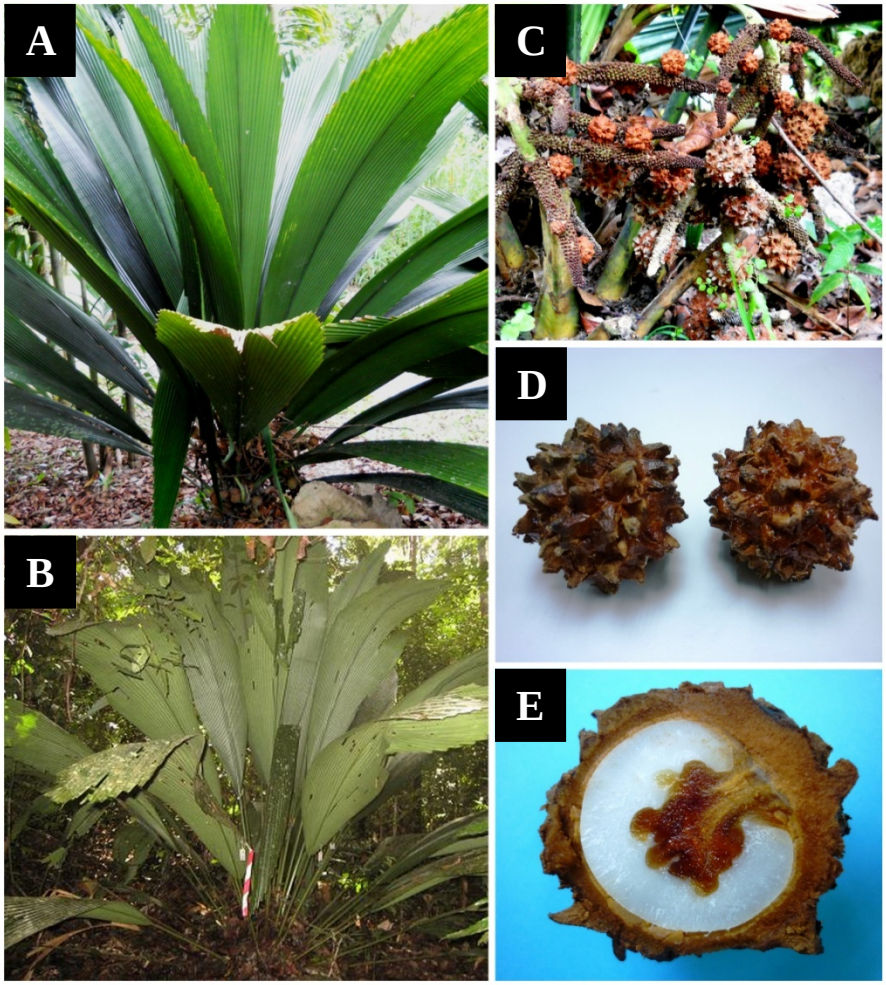

Figure 2.1. *Johannesteijsmannia lanceolata*: from Kepong Botanical Garden (A); from Angsi Reserve Forest (B); infructescences (C); fruits (D); longitudinal section of fruit (E).

In 2006, Rozainah and Sinniah carried out a study of flowering event of *J. lanceolata* in Angsi Forest Reserve in Negeri Sembilan. In a total period of 18 months observation, they found that the flowering phase from the appearance of inflorescence until the fall of mature ripe fruit last between 8.7 to 9.4 months. Overall, the species produced very small number of mature fruits in a year. Only around 0.4 % of potential fruit sets will eventually produced mature seeds. They could not find clumped distribution in the species which might be attributed to the low production of seeds and huge size of adult palms. The adult palm can grow up to 3 meters high.

On the other hand, Chan (2009) found that natural population of *J. lanceolata* showed seasonal flowering with a long reproductive cycle from flowering to fruiting. The flowers were homogamous with stingless bees (*Trigona* sp.) and small flies (Phoridae and Cevidomyiidae) as the potential pollinators. She observed that the species possessed a low fruit set, which could be attributed to the facultative selfing practice of this palm.

Look (2007) also conducted a population studies on *Johannesteijsmannia* spp. using amplified fragment length polymorphism, AFLP. Her results suggested that there is no hybridization between *J. altifrons*, *J. lanceolata* and *J. magnifica* at Sungai Lalang Forest Reserve. She suggested that *Johannesteijsmannia* spp. may exhibit outcrossing mating system based on the low genetic variation between populations. She believed that there is gene flow among populations of each species and no inbreeding occurrence in the genus. Nevertheless, her average sample size was 8 in number which are not substantial to draw a conclusion. Thus, further investigation using microsatellite markers in a larger sample size can be implemented to provide more comprehensive analysis in order to resolve all the uncertainties.

2.3 Genetic markers

Genetic markers can be divided into morphological markers, molecular markers and biochemical markers. Generally, an ideal genetic marker should possess certain criteria such as polymorphic, multiallelic, codominant, non-epistatic, neutral, and insensitive to environment. However, both morphological and biochemical markers hardly meet these criteria. Morphological markers are dorminant in general, inadequately polymorphic, interact with other traits and easily influenced by the surrounding condition. Biochemical markers such as allozymes are heterozygote deficiencies because DNA variation is hardly detected at the protein level. In addition, allozyme is also tissue specific and depends on developmental expression as well as influenced by environmental factors which caused difficulty in detection. Therefore, DNA based molecular markers are more preferable than morphological and biochemical markers in most population study of target species (Vienne, 2003).

2.3.1 DNA-based molecular markers

Introduction of molecular markers over the last two decades have changed the progression of biological sciences. Several different techniques have emerged to elucidate genetic variation owing to the rapid expansion of molecular genetics field (Parker et al., 1998). Examples of the DNA-based molecular markers are restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), microsatellite, single nucleotide polymorphism (SNP) and expressed sequences tagged (EST). These markers differ in the respect of genomic abundance, level of polymorphism, locus specificity, reproducibility, development cost, and amenable to automation (Table 2.1).

Table 2.1. Overview of the relevant characteristics of molecular markers adapted from Ng (2005).

	Allozyme	**RAPD**	**RFLP**	**AFLP**	**Microsatellite**	**SNP**
Abundance	low	high	high	high	high	high
Polymorphism	low	moderate	moderate	moderate	high	high
Specificity	yes	no	yes	no	yes	yes
Reproducibility	very high	low	high	moderate	high	high
Dominance	codominant	dominant	codominant (nDNA),haploid (cytoplasmic DNA)	dominant	codominant	codominant
Mode of inheritance	bi-parental	bi-parental	bi-parental (nDNA), uniparental (cytoplasmic DNA)	bi-parental	bi-parental	bi-parental
Number of loci	depends on the no. of enzyme genes	depends on genome size & nucleotide polymorphism	depends on restriction size	depends on restriction size	genome size & no. of SSR	depends on genome size & nucleotide polymorphism
Null alleles	rare	not applicable	extremely rare	not applicable	present	present
Transferability	genera-family	species	genera	species	species-genera	not applicable
DNA amount per assay	0.5 g	2-10 ng	2-10 mg	0.1-1 µg	5-20 ng	5-20 ng
Development cost	low	low	moderate	moderate	high	high
Cost per assay	low	low	high	moderate	low	low
Amenable to automation	low	moderate	low	moderate	high	high

2.3.1.1 Restriction Fragment Length Polymorphism, RFLP

RFLP was discovered by Botstein et al. (1980) to map human genes. RFLP involves digestion of DNA with restriction enzyme which results in different fragment sizes among various individuals, species and populations. In the past, fragments were separated by Southern Blotting whereby the digested DNA were separated by agarose gel electrophoresis and transferred to a membrane for final visualization of the probe-hybridized fragments. Nowadays, this method is more efficient when it is implemented using PCR amplification compared to Southern Blotting. RFLP is used due to its moderate polymorphism, high genomic abundancy, codominant inheritance and high reproducibility. However, RFLP requires high expense, time consuming and laborious methods. On top of it, a large amount of DNA is required for DNA digestion to generate an entire DNA fingerprinting data (Spooner et al., 2005).

2.3.1.2 Random Amplified polymorphic DNA, RAPD

RAPD is the DNA fragment amplified by PCR using short non-specific synthetic primers (around 10 bp). The short primers and low annealing temperature condition have enhanced the likelihood of amplifying anonymous region of nuclear DNA which gives rise to multiple products representing different loci (Williams et al., 1990). RAPD is used because it becomes easier and faster to assay with the implementation of PCR. Furthermore, prior knowledge of DNA sequences is unnecessary since the primers are commercially available. However, RAPD is less commonly used in fisheries research (Wirgin and Waldman, 1994) owing to the dominant inheritance of RAPD which is unable to distinguish between homozygous dominant and heterozygous. Moreover, low reproducibility is also one of the major limitations of this marker (Vignal et al., 2002).

2.3.1.3 Amplified Fragment Length Polymorphism, AFLP

AFLP is the combination of both RFLP and RAPD. It involves digestion of genomic DNA with restriction enzymes, followed by ligation of adapters to the digestion products and amplified by PCR selectively (Spooner et al., 2005). Its polymorphism is detected by the differences in length of amplified fragments via deletions, insertions, and primer site base substitutions. AFLP is used due to its high genomic abundance, moderate polymorphic and considerable reproducibility. Like RAPD, it is dominantly inherited and it does not require prior knowledge of DNA sequences. The main weakness of AFLP is the requirement of the use of automated gene sequencers for electrophoretic analysis of fluorescent labels (Liu and Cordes, 2004).

2.3.1.4 Single Nucleotide Polymorphism, SNP

SNP is the polymorphism caused by point mutation that generates variation in a single nucleotide position. It is highly polymorphic, codominantly inherited and very amenable to automation. Prior knowledge of the DNA sequences is critical in designing specific primers or probes. One major weakness of SNP is the low heterozygosity values. Generally, most SNP is restricted to one of two alleles (transition in either pyrimidines C/T or purines A/G) and thus known as bi-allelic. Apparently, SNP is less in polymorphic information content (PIC) value compared with multi-allele microsatellites (Liu and Cordes, 2004). SNP analysis might be useful in cultivar discrimination of crops when the polymorphism was difficult to assess i.e. cultivated tomato (Spooner et al., 2005).

2.3.1.5 Expressed Sequences Tagged, EST

EST is short (200-800 bp), unedited, single-pass sequences generated from arbitrary sequencing of cDNA clones (Adams et al., 1991). Availability of EST allows gene discovery, assists in gene structure identification, complements genome annotation, generates possible alternative transcripts, aids single nucleotide polymorphism (SNP) characterization and facilitates proteome analysis (Dong et al., 2005; Jongeneel, 2000; Rudd, 2003). Their sequences information has been used for the screening of microsatellite sequences (EST-SSR) or single nucleotide polymorphisms (EST-SNP) which is associated with the transcribed regions of the genome. EST-SSR was successfully applied in the characterization of accessions of wheat (Eujayl et al., 2002) and barley (Thiel et al., 2003). Meanwhile, EST-SNP was also implemented in the functional diversity studies of maize (Rafalski, 2002).

2.3.1.6 Microsatellites

Microsatellites, which are the repetitive DNA ,widely intersperse through out the genomes of both eukaryotes (Morgante et al., 2002) and prokaryotes (Gur-Arie et al., 2000). They are consisted of multiple copies of tandemly repeat motifs of 1-6 nucleotides and flanked by highly conserved regions (Chambers and MacAvoy, 2000). They can be present in both non-coding and coding regions. The term of microsatellite was introduced along with several terms such as simple sequence repeat (SSR), short tandem repeat (STR), simple sequence length polymorphism (SSLP) and variable number of tandem repeat (VNTR) (Rakoczy-Trojanowska and Bolibok, 2004). Length variation of microsatellites is analyzed by PCR with a locus specific primer pairs designed from the highly conserved flanking regions. Their variability features have made them one of the most popular genetic markers in population genetics studies. Therefore, microsatellite marker can be the ideal tool for the assessment of population structure and genetic diversity of *J. lanceolata*.

2.3.1.6.1 Classification of microsatellites

Microsatellites can be classified as perfect, imperfect, interrupted or composite based on the type of repeats (Oliveira et al., 2006). In a perfect microsatellite (e.g. AGAGAGAGAGAGAG), the repeat sequence is not interrupted by any base not belonging to the motif whereas imperfect microsatellite, there is an intrusion of different bases between the repeated motifs (e.g. AGAGAGACAGAGAG). For interrupted microsatellite, there are some interrupting bases within the repeated sequence (e.g. AGAGAGCGTATAGAGAG). In a composite microsatellite, the sequence comprised of two distinctive repeats adjacent to each other (e.g. AGAGAGAGTGTGTGTG).

2.3.1.6.2 Mutation mechanisms of microsatellites

Since the past decade, microsatellites have been extensively employed in most of the genetics studies. The molecular mechanisms account for microsatellite variation have not well been understood, although it has been known that the mutation rate of microsatellites was ranging from 10^{-2} to 10^{-6} nucleotides per locus per generation which is much higher than other part of the genome (Sia et al., 2000). Several mechanisms have been proposed to explain the high mutation rate of microsatellites such as: (1) errors during recombination, (2) unequal crossing-over and (3) polymerase slippage during DNA replication or repair (Oliveira et al., 2006).

The predominant mutation mechanism of microsatellites is most likely due to DNA replication slippage (Tautz and Schlotterer, 1994). During DNA replication, the two DNA strands can dissociate from each other and realign out of register. This can result in unpaired repeat loops. Generally, the loops are corrected by the mismatch repair system but somehow there is a small fraction which is neglected and resulting in gains or losses of one or more repeat units (Eisen, 1999). This phenomenon can also be called as slipped-strand

mispairing (Dettman and Taylor, 2004). Figure 2.2 illustrated the model for slippage during DNA replication.

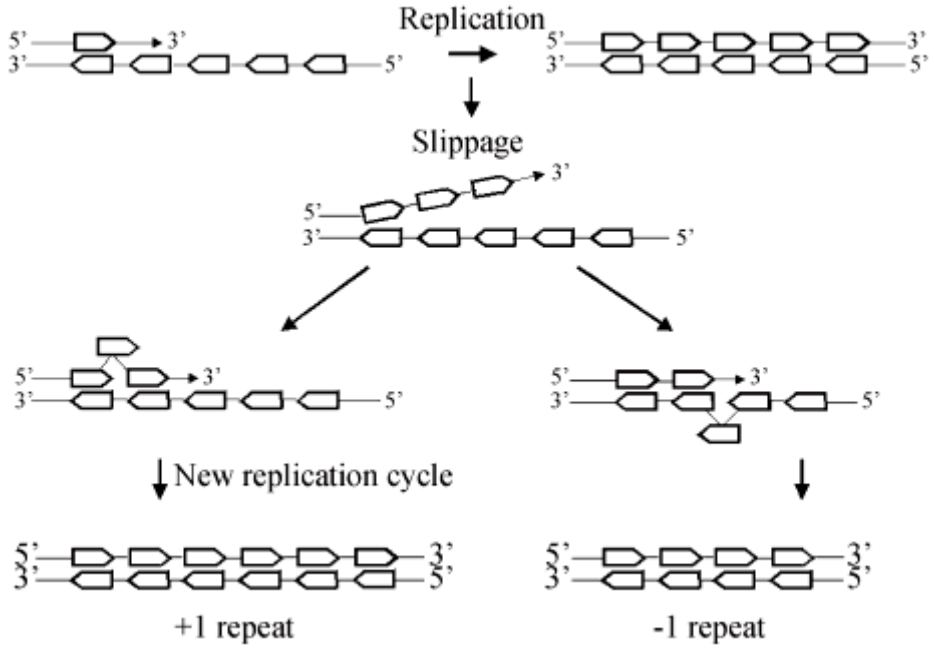

Figure 2.2. Model of microsatellites mutation by DNA replication slippage adapted from Oliveira et al. (2006). The original DNA molecules are 5 repeats of motif, denoted by a box. During DNA replication, slippage occurs and results in formation of new alleles with 6 and 4 repeats.

2.4 Applications of microsatellite markers to plant population analysis

In plant population genetics studies, knowledge of the genetic variability between and within natural population is vital to postulate competent strategies for *ex-situ* and *in-situ* germplasm conservation. In this case, microsatellites are well recommended to estimate the genetic population parameters such as (1) population structure, (2) parentage and paternity analysis and (3) gene flow.

2.4.1 Population structure assessment

Genetic structure of a population can be assessed with microsatellite markers because they are highly polymorphic compared with other molecular markers. Microsatellites have large number of alleles and thus few numbers of microsatellites are adequate to conduct the population analysis. High mutation rate of microsatellite loci also allowed the estimation of large and widespread population (Oliveira et al., 2006).

There are two measures to assess population structure namely Wright's *F*-statistics and R_{ST} statistics (Slatkin, 1995). Wright's inbreeding coefficient (F_{ST}) is very useful in discriminating rare alleles although sometimes it is overestimated from the true value. It neglects the homoplasic alleles and this may lead to overestimation of the rate of gene flow and gene introgression. Therefore, R_{ST} statistics is introduced to take into account the effect of mutation which is able to create a novel allele in estimation of the closest true value. F_{ST} is based on the infinite allele model which ignores homoplasy and considered alleles to be identical by descent whereas R_{ST} is based on a stepwise mutation model that postulates mutation altered the number of repeats through addition or deletion of one repeat unit at a fixed rate.

Krutovsky et al. (2009) conducted population structure analysis of coastal Douglas-fir using allozyme and microsatellite markers. Generally, result of both allozyme and microsatellites indicated weak differentiation among the population. Their studies showed

that allozyme and microsatellites genotypic data can be combined for joint analysis to better estimation of population structure. In Malaysia, Kumar et al. (2002a; 2002b) developed sets of microsatellite markers in order to be utilized in the assessment of mungbean's genetic structure in the wild population. They recognized genetic variation of the species and different varieties were cross-breeding for higher yield and better quality improvement.

2.4.2 Parentage and paternity analysis

In plant paternity analysis, microsatellite markers were employed not only due to their codominant inheritance, but also because they have relatively rare alleles that could be used in calculating the probability of exclusion. The probability of exclusion depends on the allele frequency of a locus, not on the genotype (Weir, 1996). Generally, 30-40 microsatellite loci are required to provide a competent estimation of relationship coefficients (Blouin, 2003). For instance, Bowers et al. (1999) used 32 microsatellite loci to identify the relationship of 300 cultivar grapes and they showed that most of the cultivars had originated from a pair of *Pinot* and *Gouais blanc* parents which widely spread in northeastern France during middle ages.

2.4.3 Gene flow

Gene flow is the exchange of genes between different populations of a species due to migration and changes the gene frequency of the recipient gene pool simultaneously. It is important in maintenance of metapopulation since it retains the genetic diversity by acting on the population structuring mechanism and against genetic drifts (Oliveira et al., 2006). It is also homogenizes allelic frequency to make population becomes more heterogenous.

In plants, gene flow can be quantified by referring to numerous loci to infer the most likely male parent of a specific offspring in paternity analysis. As soon as male parent is identified, pollen migration can be determined. Gene flow can be measured in term of

number of migrants per generation, *Nm*, where *N* is the population size and *m* is the proportion of migrants per generation. Gaggiotti et al. (1999) conducted simulation studies to estimate gene flow (*Nm*) using Wright's F_{ST} statistic which emphasized on the basis of variance of gene frequencies and R_{ST} which was quantified from the variance of length of the alleles. The studies showed that estimation of gene flow with microsatellites can lead to serious overestimation particularly if the population is very large in size (N> 5000). They suggested that during estimation of (*Nm*), R_{ST} performs better than F_{ST} when the population size is large (N= 50) and many microsatellite loci (n= 20). In contrast, F_{ST} performs better than R_{ST} when the population size small (N= 10) and number of loci is low.

2.5 Development of microsatellites

There are numerous ways to develop microsatellite markers: (1) searching through the available databases, (2) constructing and screening the genomic library, or (3) using strategies without library construction (Rakoczy-Trojanowska and Bolibok, 2004).

Development of microsatellite markers from databases even though is simple, fast and cost-effective but it is taken from expressed sequences with substantial polymorphism and solely limited to species with high commercial value or scientific interest. In library construction strategy, there are two general strategies: (1) non-enriched libraries and (2) enriched libraries. In the past, the non-enriched libraries approach was laborious and high cost. It involved hybridization of genomic libraries with microsatellite-containing probes followed by sequencing of the hybridized positive clones. On the contrary, screening microsatellite from small-insert enriched genomic library has significantly saved a lot of time and cost (Billotte et al., 2001). This approach used biotin-labeled probes either to capture by magnetic beads coated with streptavidin or fix on a nitrate filter (Edwards et al., 1996). Therefore, the enriched libraries approach was preferable to non-enriched libraries

because the eluted DNA after washing out non-hybridized DNA was highly enriched for microsatellites (Butcher et al., 2000).

By using repeat-anchored random primers (Wu et al., 1994) or hybridization of amplified RAPD products with microsatellites probes, modified RAPD approach was postulated to avoid library construction (Cifarelli et al., 1995; Richardson et al., 1995). Another approach, PCR isolation of microsatellites array (PIMA) skipped DNA fragmentation steps to cloning step has been proposed by Lunt et al. (1999). It used repeat-specific primers to detect microsatellites with assumption that RAPD fragments have higher likelihood to contain microsatellites than random genomic clones. Zane et al. (2002) proposed a fast and effective protocol named FIASCO (Fast Isolation by AFLP Sequences Containing repeats) relied on the efficient digestion-ligation reaction of AFLP and then hybridized with biotinylated probes, followed by selective capturing of microsatellites with streptavidin-coated beads.

2.6 Cross-species amplification of microsatellites

Cross-species amplification or transferability can be defined as an incidence where the primer pairs designed of a particular species can be used for other closely related species as well through PCR amplification. The other species can be under the same genus (Cipriani et al., 1999; Isagi and Suhandono, 1997) or even for different genera under the same family (Roa et al., 2000; White and Powell, 1997; Zucchi et al., 2002).

The transferability feature of microsatellites is attributed by the conserved microsatellites flanking regions among related species. It has reduced the cost of working on taxa with low microsatellites frequency and in microsatellites which were hard to be isolated. In general, the rate for successful transferability is inversely related with the phylogenetic distance of the species. In plants, successful transferability of microsatellites has been observed in cultivars, subspecies and related species developed from *Eucalyptus*

species (Brondani et al., 1998) to *Eugenia dysenterica*. Besides, cross species amplification of microsatellites was also observed in *Koompassia* malaccensis (Leguminosae), an important tropical timber species (Lee et al., 2006a). The successful cross-amplification might be associated within the genus and not in subfamily level.

However, low transferability has been described in amphibian genera *Triturus* (Garner et al., 2003) and *Rana* (Primmer and Merilä, 2002). Perhaps this is due to the very large genome of amphibians. Through the studies, phylogenetic distance is apparently the predominant factor in successful transfer yet other factors such as the size and complexity of the genome also affect the transferability.

CHAPTER 3

MATERIALS AND METHODS

CHAPTER 3

MATERIALS AND METHODS

3.1 Leaf samples collection

Young leaves of a total of 36 individuals of *J. lanceolata* were collected from Angsi Forest Reserve (2° 43′ 4.50″ N and 102° 4′ 6.30″ E), Negeri Sembilan, Malaysia. Subpinnately rib of the leaves were removed and then cut into small pieces around 1 cm to be kept at -80 ºC until further processing.

3.2 Total genomic DNA extraction

A Cetyl Trimethyl Ammonium Bromide (CTAB**)** plant DNA extraction technique modified by Murray and Thompson (1980) was used to extract total genomic DNA from the plant leaves. Approximately 3–5 grams of fresh leaf was frozen in liquid nitrogen prior to being ground. The grindate was then poured into a 50 mL falcon tube containing preheat (65 ºC) 20 mL CTAB extraction buffer (2 % [w/v] CTAB, 20 mM EDTA, 100 mM Tris-Cl pH 8.0, 1.4 M NaCl, 0.2 % β-mercaptoethanol, 1 % PVP-40) and incubated at 65 ºC for 30 minutes with occasional mixing. The mixture was then allowed to cool down to room temperature. An equal volume of chloroform-isoamyl alcohol (24:1) was added and mixed gently for 15 minutes. After centrifugation at 3,000 rpm for 10 minutes, the top aqueous layer was transferred into a new 50 mL falcon tube. Two-third volume of cold (-20 ºC) isopropanol was added and mixed gently to precipitate nuclei acids. After an hour of incubation, DNA was precipitated by centrifugation at 3,000 rpm for 10 minutes. The supernatant was then discarded and DNA pellet transferred into a 2.0 mL tube containing 1 mL of wash buffer (76 % ethanol, 10 mM ammonium acetate) and kept for at least an hour at -20 ºC prior to centrifugation at 12,000 rpm for 10 minutes, 4 ºC. Supernatant was

discarded carefully and DNA pellet was left to dry at 37 ºC. The DNA pellet was dissolved in 500 µL of TE buffer (50 mM Tris-Cl, pH 8.0, 20 mM EDTA).

The DNA sample was further purified using Qiagen DNeasy Plant Mini Kit (Qiagen GmbH, Germany). DNA concentration were quantified through electrophoresis of 0.85 % agarose gel containing 0.5 µg/mL ethidium bromide in 1X TAE buffer (40 mM Tris-acetate, pH 8.0, 1 mM EDTA) in comparison with a series of calf thymus DNA concentration markers (Boehringer Mannheim, WestGermany) as standards. DNA purity was checked by Nanodrop 2000 spectrophotometer (Thermo Scientific, Wilmington) in term of the ratio absorbance at 260 nm and 280 nm wavelength (OD_{260}/OD_{280}).

3.3 Construction of CT-enriched library

3.3.1 Digestion of genomic DNA with restriction enzyme

Approximately 5 µg of genomic DNA was digested with *Nde* II (Promega, USA). Preparation of reaction as shown in Table 3.1 was done on ice and followed by incubation at 37 ºC for 2 hours.

Table 3.1. Preparation for DNA digestion with *Nde* II.

Components	Volume per reaction (µL)	Final concentration
DNA (400 ng/µL)	12.5	5 µg
RE buffer 10X	2.0	
BSA (10 µg/µL)	0.2	
dH_2O	3.8	
Nde II (10 U/µL)	1.5	15 Units
Total	20.0	

The digested genomic DNA was subjected to 1.6 % agarose gel electrophoresis for size-selection. DNA fragments ranging from 300-1,000 bp were recovered and purified using Qiagen MinElute Gel extraction Kit (Qiagen GmbH, Germany) and eluted in 15 µL of EB buffer. DNA concentration was checked using Nanodrop 2000 spectrophotometer (Thermo Scientific, Wilmington).

3.3.2 Ligation of DNA fragments with cassettes

Approximately 500 ng of digested DNA was ligated with *Sau*3A1 cassettes (Takara, Japan) using DNA Ligation Kit Ver.2.1 (Takara, Japan). The mixture as shown in Table 3.2 was incubated at 16 ºC for 2 hours followed by heat inactivation of T4 DNA ligase at 70 ºC for 10 minutes. The DNA was precipitated by using ethanol, sodium acetate and Dr. Gentle (Takara, Japan) and then dissolved in 15 µL of dH_2O.

Table 3.2. Ligation set up of DNA fragments with *Sau*3A1 cassettes.

Components	Volume per reaction (µL)
Template DNA (~500 ng)	5
*Sau*3A1 cassettes (1.0 pmol/µL)	5
Solution II	10
Solution I (Ligase)	20
Total	40

The nicks between DNA fragments and *Sau*3A1 cassettes were repaired after ligation using DNA polymerase I (Takara, Japan) as shown in Table 3.3. The mixture was then incubated at 16 ºC for 2 hours followed by heating at 70 ºC for 10 minutes.

Table 3.3. Repair of nicks between DNA fragments and cassettes.

Components	Volume per reaction (μL)
DNA	15.0
10X buffer	3.0
dNTP (10 mM)	7.5
DNA Polymerase I (4 U/μL)	1.0
dH_2O	3.5
Total	30.0

3.3.3 PCR amplification to confirm ligation

PCR amplification as shown in Table 3.4 was carried out to check successive ligation of the *Sau*3A1 cassette to the genomic DNA fragments in the condition of initial denaturation at 95 ºC (3 min), 35 cycles of 95 ºC (1 min), 55 ºC (1 min), 72 ºC (2 min) and followed by final extension at 72 ºC (2 min).

Table 3.4. PCR reaction mixture.

Components	Volume per reaction (µL)
dH_2O	6.63
10X buffer	1.25
$MgCl_2$ (25 mM)	0.75
dNTP (2 mM)	1.25
Taq (5 U/µL)	0.12
Primer C1 (10 pmol/µL)	1.50
DNA ligated to cassettes	1.00
Total	12.5

3.3.4 Selective hybridization using biotinylated $(CT)_{15}$

The concentration of cassette-ligated genomic DNA was checked using Nanodrop 2000 spectrophotometer (Thermo Scientific, Wilmington) prior to hybridization reaction as shown in Table 3.5. The hybridization mixture was then heated to 95 ºC for 15 minutes to denature cassette-ligated genomic DNA and hybridized at 70 ºC overnight.

Table 3.5. Hybridization of cassette-ligated genomic DNA and biotinylated $(CT)_{15}$.

Components	Volume per reaction (µL)
Cassette-ligated genomic DNA (~100 ng)	5
Biotinylated $(CT)_{15}$ (1 µM)	2
Hybridization solution (12X SSC, 0.1 % SDS)	50
dH_2O	43
Total	100

A volume of 300 µL of Streptavidin MagnaSphere Paramagnetic Particles (Promega, USA) beads was rinsed with 150 µL binding and washing buffer (10 mM Tris [pH 7.5], 1 mM EDTA, 1 M NaCl). After rinsing, the hybridization mixture was transferred to beads in 150 µL of binding and washing buffer. The tubes were then sealed with parafilm and agitated at 43 ºC for 2 hours in rotary oven. Hybridization solution was discarded after placed the tubes in magnetic stand for 1 minute. The remaining beads were washed stringently using two types of washing buffer (2X SSC, 0.1 % SDS [at room temperature] and 1X SCC, 0.1 % SDS at 45 ºC & 60 ºC). A volume of 60 µL of 95 ºC TE buffer was added to the beads and heated to 95 ºC for 10 minutes in order to elute genomic DNA bound to biotinylated oligonucleotides. The DNA was precipitated by using ethanol, sodium acetate and Dr. Gentle (Takara, Japan) and then dissolved in 20 µL of dH_2O. The DNA fragments were then amplified by PCR reaction and precipitated prior to digestion.

3.3.5 Digestion of cassettes

Cassettes of amplified CT-enriched DNA were removed by digestion with *Nde* II. The digestion mixture as in Table 3.6 was prepared and incubated at 37 ºC for at least 2 hours. The mixture was then subjected to 1.6 % agarose gel electrophoresis. The amplified CT-enriched DNA without cassettes was eluted using Qiagen MinElute Gel extraction Kit (Qiagen GmbH, Germany) and its concentration was checked using Nanodrop 2000 spectrophotometer (Thermo Scientific, Wilmington).

Table 3.6. Digestion of cassettes mixture.

Components	Volume per reaction (µL)
Amplified CT-enriched DNA	60.0
RE buffer 10X	8.0
Acetylated BSA (10 µg/µL)	0.8
dH_2O	3.2
Nde II (10 U/µL)	8.0
Total	80.0

3.3.6 Ligation of CT-enriched DNA into plasmid

The CT-enriched DNA was ligated into pUC 118 *Bam* H1/BAP vector (Takara, Japan). Ligation mixture as shown in Table 3.7 was prepared and then incubated at 16 ºC for 2 hours prior to PCR amplification with M13 primers to check the ligation.

Table 3.7. Ligation of enriched DNA into plasmid.

Components	Volume per reaction (μL)
CT-enriched DNA (50 ng/ μL)	1.0
Plasmid (0.1 μg/μL)	0.5
TE buffer	3.5
Solution I (Ligase)	5.0
Total	10.0

3.3.7 Transformation of recombinant plasmids into competent cells

Qiagen PCR Cloning*plus* Kit (Qiagen GmbH, Germany) was used for the transformation of plasmids into Qiagen EZ Competent Cells. The transformation mixture was plated on LB agar containing ampicillin (100 μg/mL), IPTG (50 μM), X-gal (80 μg/mL) and incubated at 37 ºC overnight. Insert-containing clones were selected by blue/white screening.

3.4 Amplification of circular DNA

Transformed colonies were picked for amplification of circular DNA by using TempliPhi DNA Sequencing Template Amplification Kits (Amersham Biosciences, USA).

3.4.1 DNA sequencing and analysis of sequence data

The circular DNA was subsequently sequenced using Applied Biosystems 3130*xl* Genetic Analyzer. The reaction mixture consisted of 0.5 μL of Bigdye Terminator Mastermix (Applied Biosystems, USA), 2.0 μL of template DNA, 0.5 μL of 10 μM M13 forward primer, 1.0 μL of 5X sequencing buffer in a total volume of 10 μL. The cycle sequencing was programmed at initial denaturation 96 ºC (1 min) and 50 cycles of 96 ºC (10 sec), 50 ºC (5 sec), 60 ºC (4 min). The PCR products were purified through ethanol precipitation prior to sequencing.

The sequence data were analysed using Sequencher 4.9 (Gene Codes Corporation, USA) to trim ends and vector sequences. The 'clean' sequences were then clustered using Cd-hit program (Li and Godzik, 2006) based on sequence similarity. Microsatellites were identified using MISA-MIcroSAtellite identification tool based on the unique sequences from the clustering (Thiel et al., 2003).

3.5 Primer design

After identification of unique sequences containing microsatellites, forward and reverse primers flanking the microsatellite region were designed using OLIGO 6 software (Molecular Biology Insight, USA) and then synthesized. Some criteria were applied when designing primer pairs:

i. Primer length ranged from 18-24 bp for specificity
ii. Melting temperature (T_m) of primers lower than 65 ºC to avoid secondary annealing
iii. CG content of 40-60 %
iv. GC clamp to promote specific bonding due to stronger bonding of G & C bases
v. Sequences with high intensity of secondary structure formation such as hairpin loop and primer dimer were avoided
vi. Sequences with consecutive repeats were avoided
vii. T_m difference between primer pairs not exceeding 5 ºC
viii. Low internal stability at 3'-end i.e. less stable than -8.6 ± 0.6 kcal/mol to reduce false priming

3.6 Primer screening and PCR amplification

Four individuals of *J. lanceolata* used for primer screening were randomly chosen from the 36 individuals collected from Angsi Forest Reserve. PCR amplifications were carried out in a reaction volume of 10 µL with approximately 10 ng of template DNA, KCl (50 mM), Tris-Cl (20 mM, pH 8.0), $MgCl_2$ (1.5 mM), each primer (0.2 µM), dNTP (0.2 µM) and *Taq* DNA polymerase (0.5 Unit). The PCR thermal profile was: initial denaturation at 94 ºC (5 min), 35 cycles of 94 ºC (30 sec), 45 ºC (30 sec), 72 ºC (30 sec) and followed by final extension at 72 ºC (7 min). The PCR products were electrophoresed on 2 % agarose gel.

3.7 Fragment analysis

Forward primers of potential polymorphic amplification without multiple bands were labelled with fluorescent (6-FAM or HEX) for genotyping purposes. The PCR products with fluorescent were subjected to fragment analysis using Applied Biosystems 3130*xl* Genetic Analyzer. ROX 400 (Applied Biosystems, USA) was used as internal size standard in assigning allele sizes and further scored by using GeneMapper v 4.0 (Applied Biosystems, USA).

3.8 Characterization of microsatellite loci

Twenty-four individuals of *J. lanceolata* from Angsi Forest Reserve were used for polymorphism screening of the microsatellite loci. The polymorphic microsatellite loci were further characterized using CERVUS version 3.0.3 (Kalinowski et al., 2007) to generate number of alleles per locus, observed and expected heterozygosities and polymorphic information content (PIC). Exact tests for Hardy-Weinberg Equilibrium and linkage disequilibrium were checked using GenePop version 4.0.10 (Rousset, 2008). Micro-checker (Van Oosterhout et al., 2004) was further used to aid the identification of genotyping errors and to confirm the presence of null alleles in any loci.

3.9 Analysis of inheritance

Segregation of the alleles in an open-pollinated half-sib family was examined to verify the Mendelian mode of inheritance for the polymorphic microsatellites. Fifteen seeds were collected from a mother tree and DNA was extracted from the embryo (Figure 3.1) using Qiagen DNeasy Plant Mini Kit as described in Section 3.2. PCR amplification and fragment analysis was subsequently performed following the same protocol used for leaves.

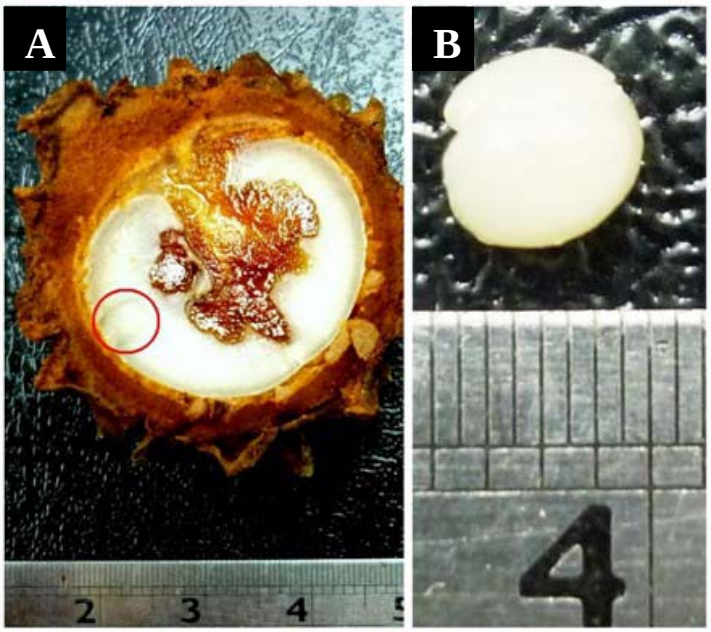

Figure 3.1. A: Longitudinal section of fruit of *J. lanceolata* with embryo in the red circle. B: Embryo.

3.10 Cross-species amplification

Sampling of three other species of *Johannesteijsmannia*: *J. altifrons* from Endau-Rompin National Park, Johor (2° 26′ 20.09″ N and 103° 16′ 22.29″ E), *J. magnifica* from Jeram Toi Recreational Forest, Negeri Sembilan (2° 51′ 52.98″ N and 102° 0′ 50.34″ E) and *J. perakensis* from Kledang Sayong Forest Reserve, Perak (4° 42′ 57.24″ N and 100° 58′ 2.16″ E) was carried out to check transferability of the microsatellites. Total genomic DNA of the respective species was extracted from young leaf as described in Section 3.2. Four individuals of each species were selected for the amplification of polymorphic microsatellite analysis.

CHAPTER 4
RESULTS

CHAPTER 4

RESULTS

4.1 Construction of CT-enriched library

A CT-enriched library of *J. lanceolata* was constructed to isolate microsatellites according to a modified method from Lee et al. (2004b). Total genomic DNA was successfully extracted from the young leaves of *J. lanceolata* through the modified CTAB extraction method (Murray and Thompson, 1980). The extracted DNA samples were relatively pure with the ratio absorbance at 260 nm and 280 nm wavelengths (OD_{260}/OD_{280}) ranging from 1.7 to 2.0. DNA yields determined were rather satisfactory with about 33.4–148.6 µg produced from approximately 5 grams of fresh leaves. DNA concentration was estimated through electrophoresis of 0.85 % agarose gel in comparison with a series of calf thymus DNA concentration markers as standards. Figure 4.1 shows the electrophoresis of the total genomic DNA extracted from two individuals of *J. lanceolata*.

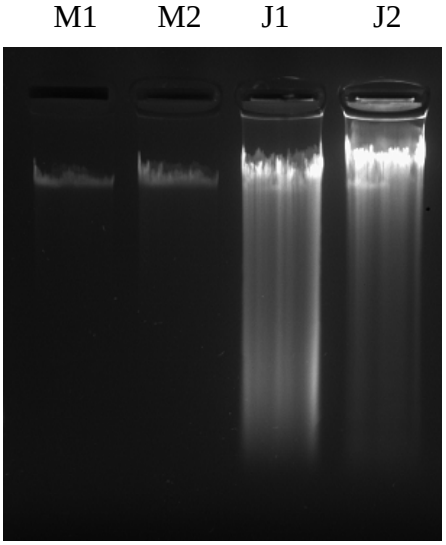

Figure 4.1. Ethidium bromide stained 0.85 % (w/v) agarose gel of extracted genomic DNA.

M1 & M2: 50 ng and 100 ng DNA concentration markers

J1 & J2: Genomic DNA extracted from two individuals of *J. lanceolata*

Genomic DNA of individual J2 was selected and further fragmentized via digestion with *Nde* II restriction enzyme. The digested genomic DNA was subjected to 1.6 % agarose gel electrophoresis for size-selection. Figure 4.2 shows the electrophoresis of the digested genomic DNA with fragments predominantly ranging from 100 bp to 1,000 bp in size.

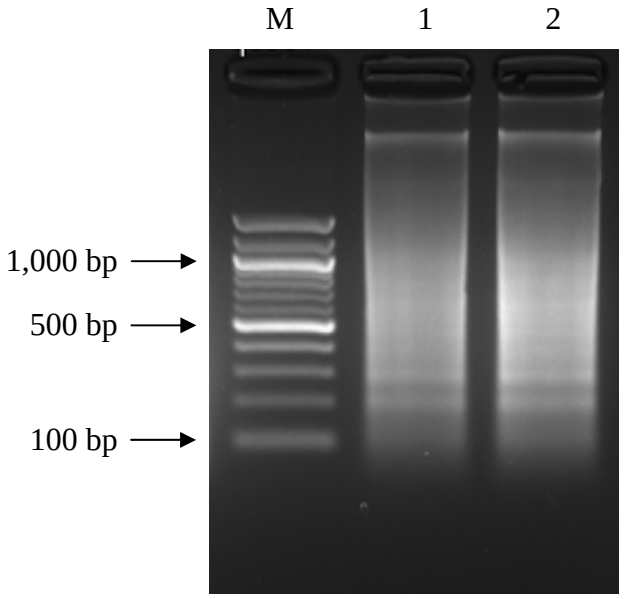

Figure 4.2. Ethidium bromide stained 1.6 % (w/v) agarose gel of digested genomic DNA with *Nde* II.

M: 100 bp marker

Lane 1 & 2: Digested genomic DNA

DNA fragments ranging from 300 bp to 1,000 bp were recovered and purified prior to the ligation with *Sau*3A1 cassettes. The nicks between DNA fragments and cassettes were repaired using DNA polymerase I. PCR amplification was then performed using C1 primer to further confirm the ligation of DNA fragments with the cassettes. Figure 4.3 shows the electrophoresis of the PCR amplicons (cassette-ligated DNAs) ranging from size of 300 bp to 1,000 bp.

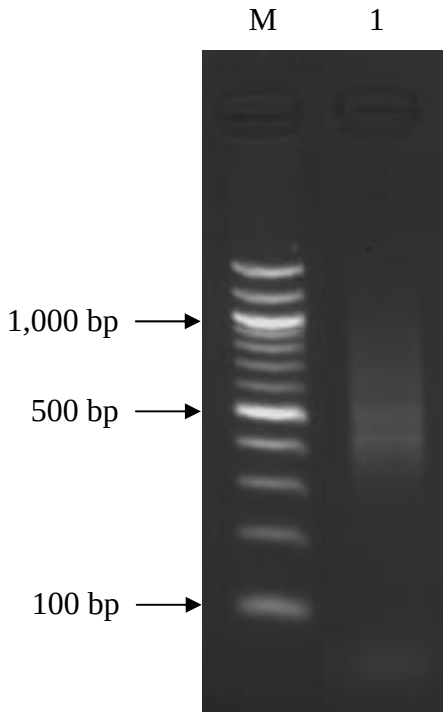

Figure 4.3. Ethidium bromide stained 1.6 % (w/v) agarose gel of PCR amplicons.

M: 100 bp marker

Lane 1: PCR amplicons that confirmed the ligation of *Sau*3A1 cassettes with the DNA fragments ranging from size of 300 bp to 1,000 bp

The cassette-ligated DNAs were then hybridized to biotinylated $(CT)_{15}$ probe and subsequently bound to streptavidin coated magnetic beads. A magnetic stand was placed beside the tube to pull these hybrids to the wall of tube. The non-hybridized DNAs were then removed by two stringent washing steps. The enriched DNA fragments were eluted and amplified by PCR. The amplified enriched DNA molecules were digested with *Nde* II to remove the cassettes. Figure 4.4 shows the electrophoresis of the enriched DNA fragments after removal cassettes. The enriched DNA fragments were ranged around 250 bp to 1,000 bp in size whereas the removed cassettes were less than 100 bp.

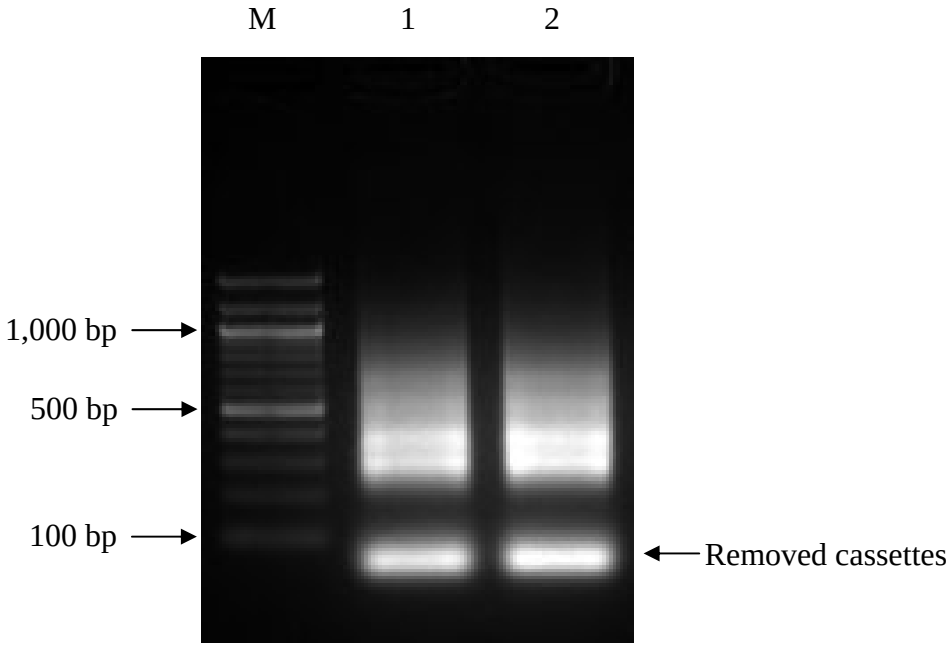

Figure 4.4. Ethidium bromide stained 1.6 % (w/v) agarose gel of enriched DNA fragments after removal of cassettes.

M: 100 bp marker

Lane 1 & 2: Size of enriched DNA fragments after removal of cassettes ranging from 250–1,000 bp

The enriched DNA fragments were then ligated into pUC118 *Bam* H1/BAP vector, transformed into Qiagen EZ competent cells and plated on LB agar. A total of 336 white colonies were selected for the amplification of circular DNA. Figure 4.5 shows the growth of transformed white colonies after overnight incubation at 37 ºC. Blue colony was not observed in this plate indicates the absence of non-ligated plasmid in the colonies.

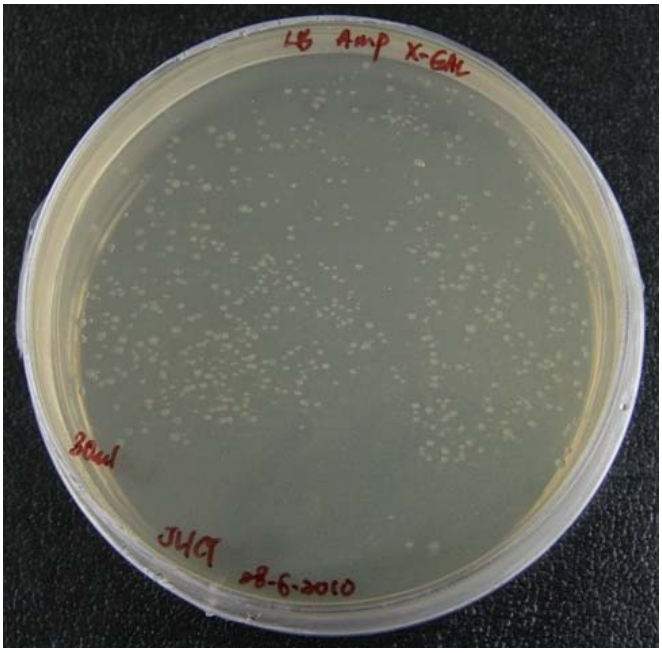

Figure 4.5. A LB agar plate containing ampicillin, IPTG and X-gal showing the growth of transformed white bacterial colonies after overnight incubation at 37 ºC.

4.2 DNA sequencing of positive clones

Circular DNA from these white clones were amplified using the Templiphi kit and sequenced using an Applied Biosystems 3130*xl* Genetic Analyzer with BigDye Terminator v3.1 Cycle Sequencing Kits and M13 forward primer. Based on the sequencing of 336 circular DNA, a total of 285 non-redundant sequences were obtained at 0.95 sequence identity threshold value, revealing 15.2 % of clone redundancy. Figure 4.6 shows some chromatograms with microsatellite sequences.

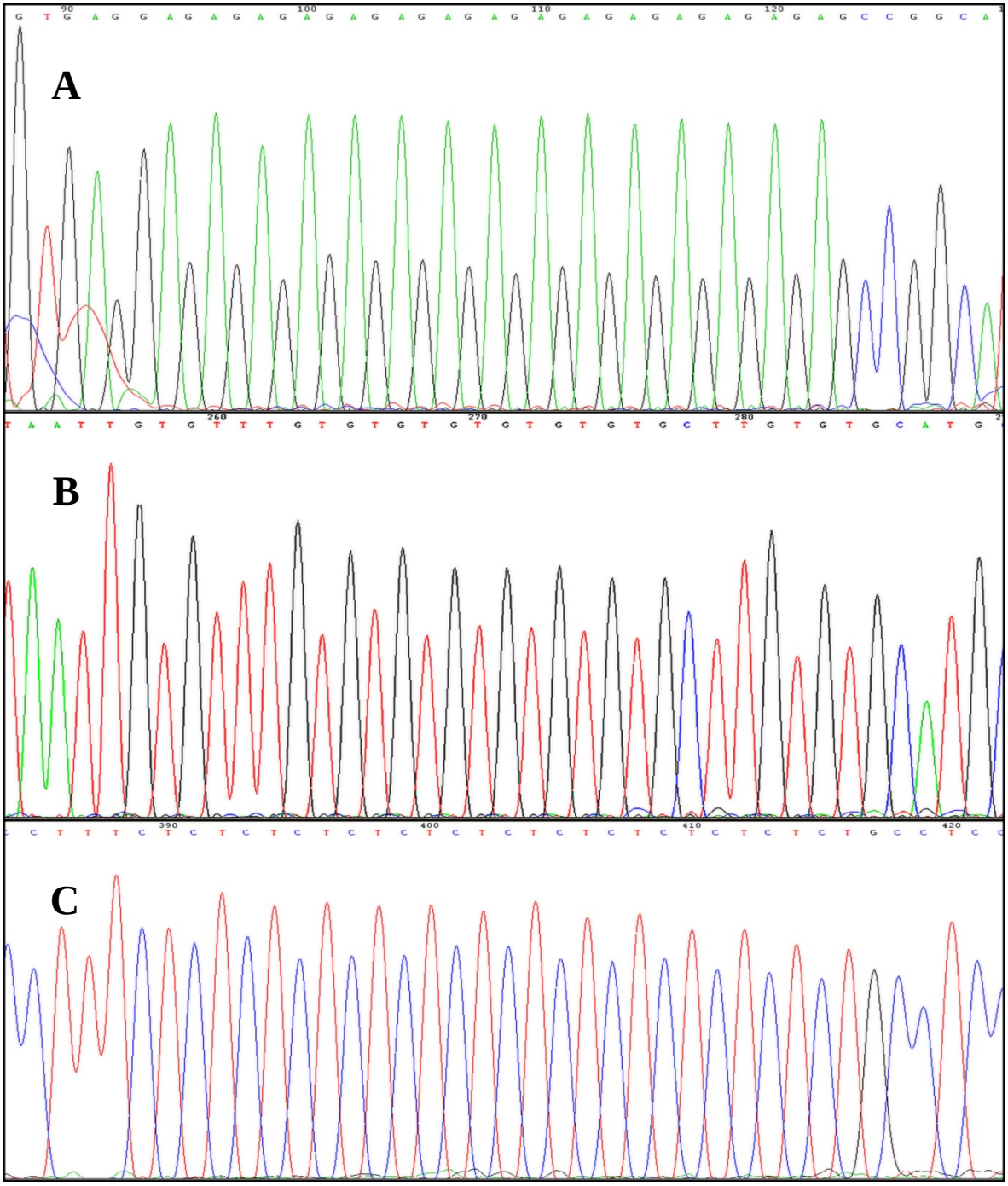

Figure 4.6. Chromatograms showing microsatellite sequences from clone A, *Jla*048: $(GA)_{15}$ repeat; clone B, *Jla*187a: $(TG)_8$ repeat and clone C, *Jla*332a: $(CT)_{14}$ repeat.

Microsatellites were identified based on the criteria (unit size/minimum number of repeats): (2/5) (3/4) (4/3) (5/3) and up to 25 maximal numbers of bases interrupting between two microsatellites in an interrupted microsatellite. A total of 353 microsatellites with different unit sizes were identified from 226 clones. Dinucleotides microsatellites showed the highest number of microsatellites among all the microsatellites (Table 4.1).

Table 4.1. Total numbers of microsatellites identified according to the respective unit sizes.

Unit size	**Number of microsatellite**
Dinucleotides	322
Trinucleotides	10
Tetranucleotides	15
Pentanucleotides	6
Total	353

The dinucleotides microsatellite can be categorized into four groups based on their repeat motifs: (AC/GT); (AG/CT); (AT/AT); (CG/CG). The AG/CT repeat motifs showed the highest frequency (78.5 %) among all the different repeat motifs found in the library (Table 4.2) as these clones were isolated from a CT enriched library.

Table 4.2. Frequency of different microsatellites repeat motifs identified from the CT-enriched genomic library of *J. lanceolata.*

Repeats	**Frequency**	**Percentage**
AC/GT	39	11.0
AG/CT	277	78.5
AT/AT	2	0.6
CG/CG	4	1.1
AAG/CTT	6	1.7
AGG/CCT	4	1.1
AAAC/GTTT	2	0.6
AAAT/ATTT	1	0.3
AAGC/CGTT	1	0.3
AAGG/CCTT	1	0.3
AATT/AATT	1	0.3
ACAG/CTGT	1	0.3
ACGC/CGTG	2	0.6
ACTC/AGTG	1	0.3
AGAT/ATCT	2	0.6
AGGG/CCCT	3	0.8
AAAAC/GTTTT	1	0.3
AAAAG/CTTTT	1	0.3
AAGAG/CTCTT	1	0.3
ACCCG/CTGGG	2	0.6
AGCCC/CGGGT	1	0.3
Total	353	100.0

4.3 Primer design

Base on the 353 microsatellites, a total of 111 forward and reverse primers flanking the microsatellite regions were designed and synthesized. Among the 111 pairs of primers, 102 (91.9 %) pairs amplify perfect microsatellites; six (5.4 %) pairs amplify interrupted microsatellites and three (2.7 %) pairs amplifying composite microsatellites.

4.4 Primer screening

Four individuals of *J. lanceolata* were used for primer screening. Total genomic DNA of *J. lanceolata* individuals were extracted using the modified CTAB extraction method. PCR products from the primer screening can be categorized into groups of: 1) specific amplicon within expected size (Figure 4.7 A); 2) specific amplicon out of expected size (Figure 4.7 B); 3) multiple amplifications (Figure 4.7 C); and 4) no amplification (Figure 4.7 D). Out of 111 primer pairs screened, 36 pairs produced specific amplicons of the expected sizes (50-350 bp) without multiple bands. These primers were further labeled with fluorescent (6-FAM or HEX) at the forward primer and further screened for polymorphism in 24 individuals of *J. lanceolata* via fragment analysis

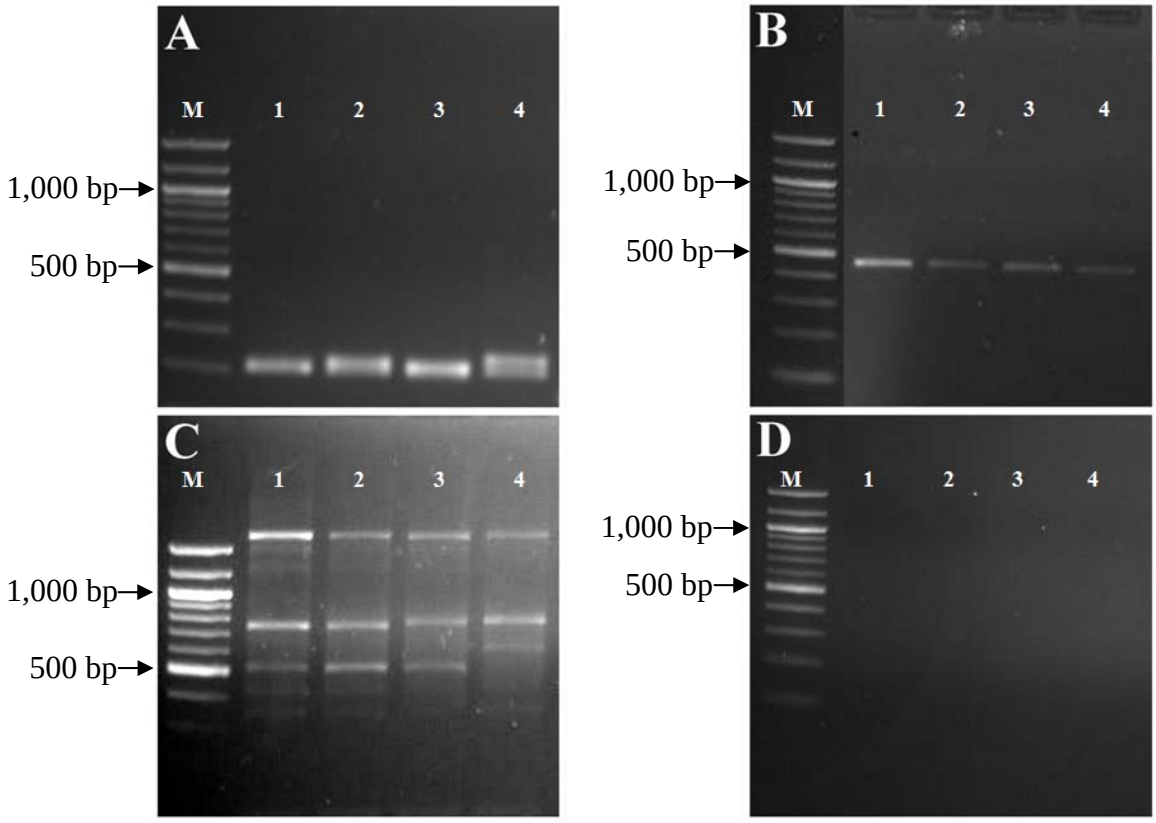

Figure 4.7. Ethidium bromide stained 2.0 % (w/v) agarose gel of PCR products generated using four primer sets.

A, *Jla*002: Specific amplicon within expected size, 206 bp

B, *Jla*070: Specific amplicon about 500 bp out of expected size, 154 bp

C, *Jla*021a: Multiple amplifications

D, *Jla*083: No amplification

M: 100 bp marker

Lane 1-4: Four different individuals of *J. lanceolata* used for primer screening

4.5 Characterization of microsatellite loci

From the polymorphism screening of the 36 primer pairs, 33 pairs exhibited polymorphism with consistent and scorable genotypes. The rest were either monomorphic (*Jla*156b) or multiple peaks with intricate pattern (*Jla*307 & *Jla*223). Figure 4.8 shows examples of polymorphic alleles amplified using particular primer pairs.

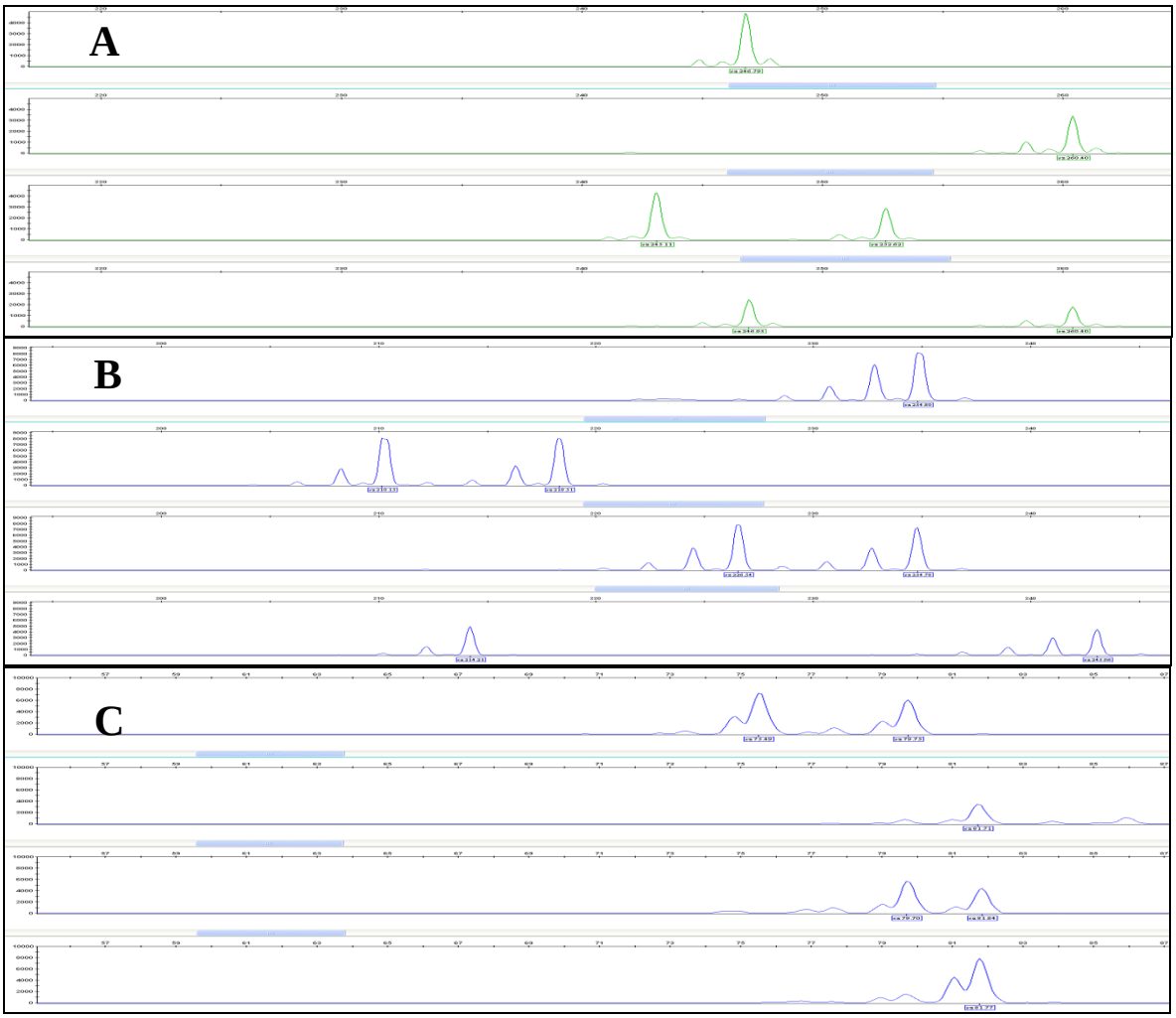

Figure 4.8. Electropherograms of polymorphic alleles amplified from four individuals of *J. lanceolata* using primer pairs A, *Jla*117; B, *Jla*112 and C, *Jla*160.

The genotypic data of these 33 polymorphic loci in 24 samples of *J. lanceolata* were analyzed using CERVUS version 3.0.3 (Kalinowski et al., 2007) and the results are given in Table 4.3. The number of alleles per locus (A) observed from the 24 samples ranged from two (*Jla*187a) to 20 (*Jla*196a) with an average of 10. The observed heterozygosity (H_O) and expected heterozygosity (H_E) ranged from 0.250 (*Jla*332a) to 1.000 (*Jla*192) and from 0.488 (*Jla*187a) to 0.957 (*Jla*196a), respectively. Polymorphic information content (PIC) ranged from 0.364 (*Jla*187a) to 0.932 (*Jla*196a) with a mean value 0.771 for all 33 loci.

Exact tests for Hardy-Weinberg Equilibrium and linkage disequilibrium were carried out using GenePop version 4.0.10 (Rousset, 2008), and Bonferroni correction for multiple tests was implemented. Significant departure from the Hardy-Weinberg equilibrium was detected in *Jla*130b and *Jla*332a with ($P < 1.515 \times 10^{-3}$). Micro-checker (Van Oosterhout et al., 2004) analysis suggested that *Jla*118 and *Jla*332a were likely to contain null alleles at 99 % confidence interval. No significant linkage disequilibrium was detected between any pair of loci ($P < 9.47 \times 10^{-5}$).

Table 4.3. Characteristics of the 33 polymorphic microsatellite loci which include locus name, repeat motifs, forward (F) and reverse (R) primer sequences, annealing temperatures (T_a), number of alleles per locus (A), allele size, observed (H_O) and expected (H_E) heterozygosities and polymorphic information content (PIC) and GenBank accession numbers. Significant departure from Hardy-Weinberg equilibrium after Bonferroni adjustment ($P < 1.515 \times 10^{-3}$) is denoted with an asterisk.

Locus	Repeat	Primer sequence (5'-3')	T_a (°C)	A	Allele size (bp)	H_O	H_E	PIC	GenBank accession no.
*Jla*002	$(GA)_{16}$	F: GGTGTGGTGCAAGGGAGTAT R: TCTCATCTACTTGGACGTCAGTGT	45	13	201; 203; 205; 207; 211; 213; 215; 217; 221; 223; 229; 231; 233	0.913	0.902	0.872	HQ613842
*Jla*046a	$(CT)_{24}$	F: TTGCCTATGTAAAATGTTAACTAA R: AGAGGGTGAGGGTTCGATT	45	19	203; 209; 211; 213; 219; 221; 223; 225; 227; 231; 233; 235; 237; 239; 241; 245; 247; 255; 257	0.864	0.911	0.881	HQ613843
*Jla*048	$(GA)_{15}$	F: TTCCCTTCTTTGCACGATTTCGAG R: GCATGGGAGGGTTGGTAGGGTTT	45	7	140; 142; 144; 148; 150; 152; 154	0.667	0.803	0.757	HQ613844
*Jla*106	$(CT)_{20}$	F: GCCCATGATTAGCTTTAACC R: TATGTTTGCCTCCCTCAGACTTGA	55	10	228; 230; 232; 234; 236; 240; 244; 246; 248; 254	0.917	0.882	0.849	HQ613845
*Jla*112	$(CT)_{21}$	F: AATGTACTTTCCCTGAGCTCTACT R: ATCACTGCCAATCTTATCACC	55	13	208; 210; 214; 216; 218; 222; 224; 226; 228; 232; 234; 236; 242	0.913	0.899	0.868	HQ613846
*Jla*117	$(GA)_{13}$	F: TTGGCAATGGAGAGAACGAGAACA R: CCAAATAATTTCAAGCGTTCCTC	45	7	243; 247; 251; 253; 257; 259; 261	0.739	0.829	0.784	HQ613847

Locus	Repeat	Primer sequence (5'-3')	T_a (°C)	A	Allele size (bp)	H_O	H_E	PIC	GenBank accession no.
*Jla*118	$(CT)_{22}$	F: CACTCCAAGAAGAACCCCGATG R: ATCTTTCTCTTTTTGCTCCGCTCC	50	11	139; 143; 149; 151; 153; 155; 159; 161; 167; 169; 171	0.636	0.877	0.842	HQ613848
*Jla*124	$(CT)_{19}$	F: AGTGGCCTTTGAATTATGTC R: CCTCCATGGCTACAACAATTAGTT	45	11	193; 197; 203; 207; 209; 211; 213; 219; 221; 271; 273	0.682	0.782	0.740	HQ613849
*Jla*130b*	$(CT)_{24}$	F: CAACTTGGCCTATGCTCAC R: AGGGACCCATTAAGTAAAGCAACT	45	9	113; 115; 117; 119; 121; 123; 129; 131; 133	0.708	0.849	0.811	HQ613850
*Jla*131	$(CT)_{28}$	F: GGGACAATAAATCATGAACAATTA R: AGACATGCACATGCACACG	45	10	77; 79; 85; 87; 89; 91; 93; 95; 97; 99	0.667	0.793	0.752	HQ613851
*Jla*160	$(CT)_{15}$	F: CCCTCCTCTCTGTCTCCCCATTTC R: AACTCCCCAAACAGCACGCAATC	45	5	76; 80; 82; 86; 90	0.542	0.621	0.531	HQ613852
*Jla*162	$(CT)_{25}$	F: CCCACAGCCCCACTGATTGATT R: TGTGTGTACAGTCCGTGCCGTCAC	45	9	179; 181; 183; 185; 187; 191; 193; 199; 201	0.792	0.815	0.771	HQ613853
*Jla*168b	$(GA)_{25}$	F: GAGAGTGCCCACTTGAGTCA R: GTCCAACAAGCATCAGACCCTTAT	45	18	180; 186; 188; 190; 192; 194; 198; 202; 204; 206; 208; 210; 212; 216; 218; 220; 226; 228	0.833	0.941	0.916	HQ613854
*Jla*174	$(CT)_{19}$	F: CAGAGGTAATGCAAAATCAACCCC R: TTATTGCAGATATGGCCCTATT	45	7	89; 93; 95; 97; 98; 103; 109	0.708	0.775	0.727	HQ613855

Locus	Repeat	Primer sequence (5'-3')	T_a (°C)	A	Allele size (bp)	H_O	H_E	PIC	GenBank accession no.
*Jla*186	$(CT)_{17}$	F: ACCAAACACGGACATTCTCAAGAT R: CGGCTCATTGGAATATGTCCT	45	11	160; 166; 168; 170; 172; 176; 178; 180; 182; 188; 190	0.833	0.845	0.811	HQ613856
*Jla*187a	$(TG)_8$	F: CCACTCCCAATTAATTAATGATTA R: GCACACGCATGCACACAAG	45	2	64; 66	0.458	0.488	0.364	HQ613857
*Jla*187b	$(GA)_{18}$	F: GTGCATGCGTGTGCGTGCG R: TTTGGTACTAATGTCTCCGGTTGA	45	7	141; 143; 145; 155; 157; 159; 183	0.739	0.853	0.813	HQ613857
*Jla*192	$(CT)_{21}$	F: TTACATGTTTTGGTATCAGGTTTT R: CAACACGAGGAATAGGCTA	45	11	138; 140; 150; 152; 158; 160; 162; 163; 164; 166; 179	1.000	0.905	0.875	HQ613858
*Jla*196a	$(CT)_{27}$	F: AAATGTACCCAAATCACAACTCAA R: TACAAGCGAGTGCACGGATA	45	20	169; 171; 173; 175; 177; 179; 185; 187; 190; 191; 192; 194; 195; 196; 197; 201; 203; 205; 207; 209	0.913	0.957	0.932	HQ613859
*Jla*198	$(GA)_{21}$	F: CATCAAAATCAAGAATCTTATGGC R: CGGATGAATAATACGTGCAAGG	45	13	251; 257; 259; 261; 263; 265; 267; 273; 275; 277; 281; 285; 291	0.958	0.902	0.872	HQ613860
*Jla*210	$(CT)_{22}$	F: TGTCCTCTACTTCCCACAACAATC R: CGCAATGTCGTGGTCTACAGT	45	12	149; 153; 159; 163; 165; 167; 169; 171; 173; 175; 187; 195	0.792	0.881	0.848	HQ613861
*Jla*236	$(CT)_{20}$	F: AGGCCTGATTCCACTGCACTAGAT R: GCATGAGAAAACGAAAGGGGTATA	45	7	75; 79; 85; 87; 89; 91; 93	0.792	0.762	0.705	HQ613862

Locus	Repeat	Primer sequence (5'-3')	T_a (°C)	A	Allele size (bp)	H_O	H_E	PIC	GenBank accession no.
*Jla*238	$(CT)_{20}$	F: CACAAGATTTTTAGCAGGTAGGAT R: ACTGGCCATCAATATTAGG	45	12	176; 178; 180; 182; 184; 186; 188; 190; 194; 196; 198; 200	0.875	0.905	0.876	HQ613863
*Jla*244	$(GA)_{12}$	F: GCTTCTTTAAATGGGGACGAC R: CCGAAAGTCAGAGCTAATGTAGGG	45	11	190; 192; 193; 194; 195; 196; 198; 204; 206; 208; 212	0.739	0.671	0.641	HQ613864
*Jla*272	$(CT)_{15}$	F: GGGCTCGTCCTTCTCAATAG R: CAAAGTTCAAAAGAAATTGAGGCA	45	9	98; 100; 104; 105; 106; 107; 108; 110; 121	0.667	0.841	0.801	HQ613865
*Jla*273a	$(GA)_{20}$	F: GCCCTTAACTAACTCGCTTTG R: CATCAGCTCCTACATCCGACGGTC	45	11	181; 183; 187; 189; 191; 193; 195; 197; 199; 201; 203	0.957	0.909	0.879	HQ613866
*Jla*282	$(GA)_{20}$	F: CCTATGACAGAGAACCCGTTCACC R: ACCATCACCTCACGGCAATAACAC	45	6	114; 116; 118; 120; 122; 124	0.833	0.809	0.763	HQ613867
*Jla*299	$(GA)_{20}$	F: TGTAGATGGTGCGGACGACT R: CGTTCCTGTCAACCTACGCTTCAA	45	8	164; 170; 172; 176; 178; 180; 188; 190	0.833	0.833	0.793	HQ613868
*Jla*305	$(GA)_{13}$	F: GAAAAGCTTCCGCCATACTAAAC R: GAGAGATTACAGCCGTCCGATTCC	45	8	78; 84; 88; 90; 94; 96; 98; 100	0.652	0.794	0.746	HQ613869
*Jla*328	$(CT)_{25}$	F: CCACACCACTTGACCGGTAAGC R: CGATTTTTGGCGAATGTAGTGATG	55	14	96; 98; 100; 104; 106; 108; 110; 114; 116; 120; 122; 124; 134; 144	0.917	0.834	0.803	HQ613870

Locus	Repeat	Primer sequence (5'-3')	T_a (°C)	A	Allele size (bp)	H_O	H_E	PIC	GenBank accession no.
*Jla*330b	$(CT)_{24}$	F: GTTTCCCACCCATCCATTAATTC R: CCAACGGTCAGTGCCTTTATATCT	45	10	115; 123; 125; 127; 129; 133; 137; 145; 151; 153	0.870	0.786	0.746	HQ613871
*Jla*332a*	$(CT)_{14}$	F: ATCCCATCGCACGAACCTT R: ATAGGAGAGTCTTTCATGCCTTTG	45	6	69; 71; 73; 75; 77; 85	0.250*	0.647	0.577	HQ613872
*Jla*332b	$(TA)_8$	F: CCCTCAAAGGCATGAAAGACTCTC R: GCCCTTCTTGACTTGTAGTACCTC	45	4	139; 140; 141; 143	0.391	0.542	0.491	HQ613872

4.6 Analysis of inheritance

All 33 loci were selected to verify the Mendelian mode of inheritance of microsatellites in an open-pollinated half-sib family consisting of 15 progenies. The results showed that the amplified fragments of all loci segregated in a codominant manner, and each of the progeny possessed at least one maternal allele, which supported the postulation of a single-locus mode of inheritance in all the 33 loci. Figure 4.9 shows the electropherograms of allele's segregation in a half-sib family of *J. lanceolata* at locus *Jla*168b. All progenies inherited at least one allele from the heterozygous maternal genotype as shown in two dotted lines.

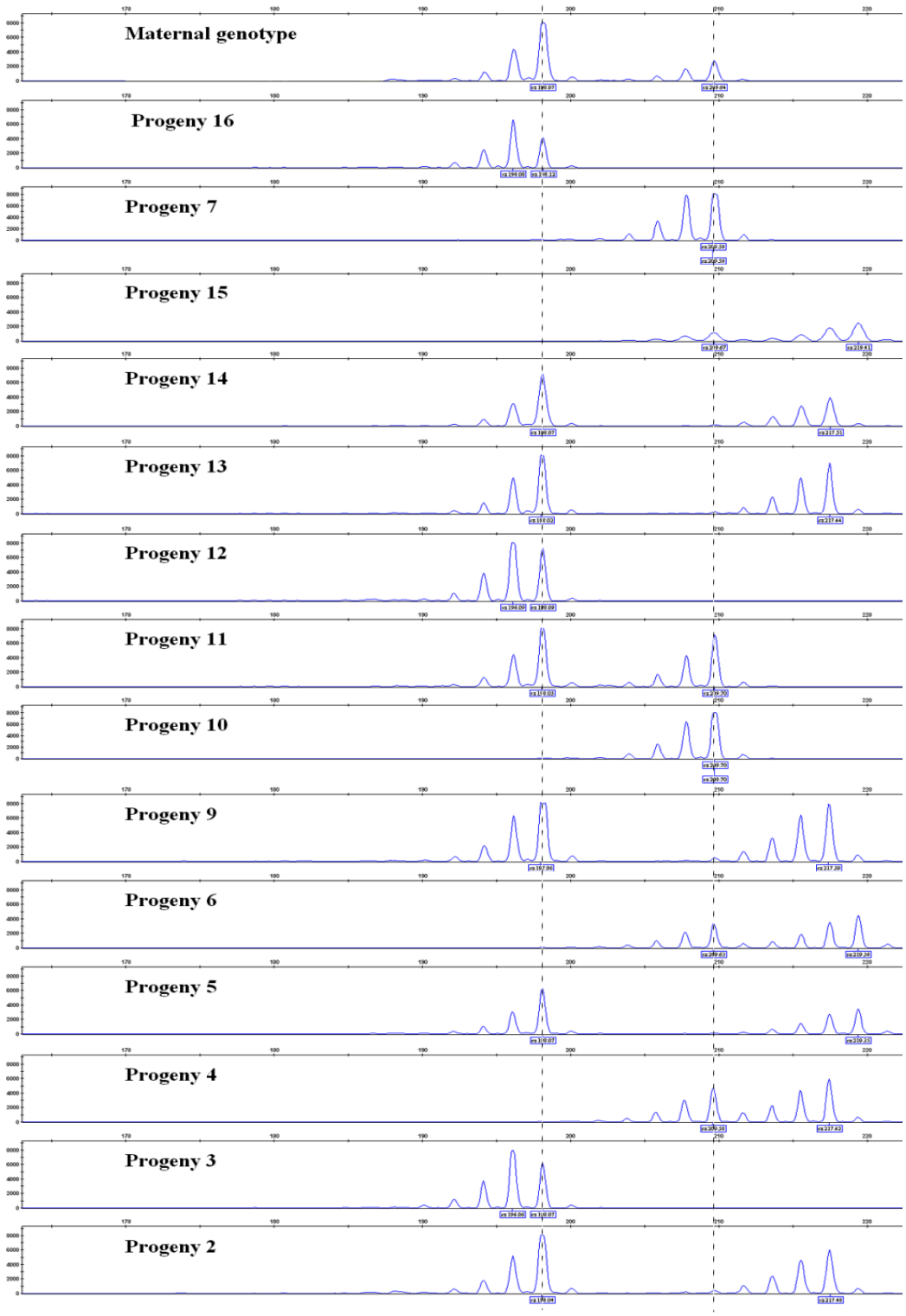

Figure 4.9. Electropherograms of allele's segregation shows in a half-sib family at locus *Jla*168b. All alleles are codominantly segregated and each of the progeny possessed at least one of the maternal alleles as shown in two dotted lines, confirming Mendelian inheritance and suggesting no seed contamination.

4.7 Cross-species amplification

Four individuals of each species, i.e., *J. altifrons, J. magnifica* and *J. perakensis,* were used to check the cross-species amplification of the microsatellite loci. In general, all 33 loci showed positive cross-species amplifications in the three *Johannesteijsmannia* species, except *Jla*124 (*J. magnifica*) and *Jla*168b (*J. magnifica* and *J. perakensis*). Amplification was considered positive when the primers produced specific PCR products of expected size without multiple bands. Figure 4.10 shows the electrophoresis of specific PCR amplicons generated from the positive cross-species amplification of locus *Jla*106 in three *Johannesteijsmannia* species.

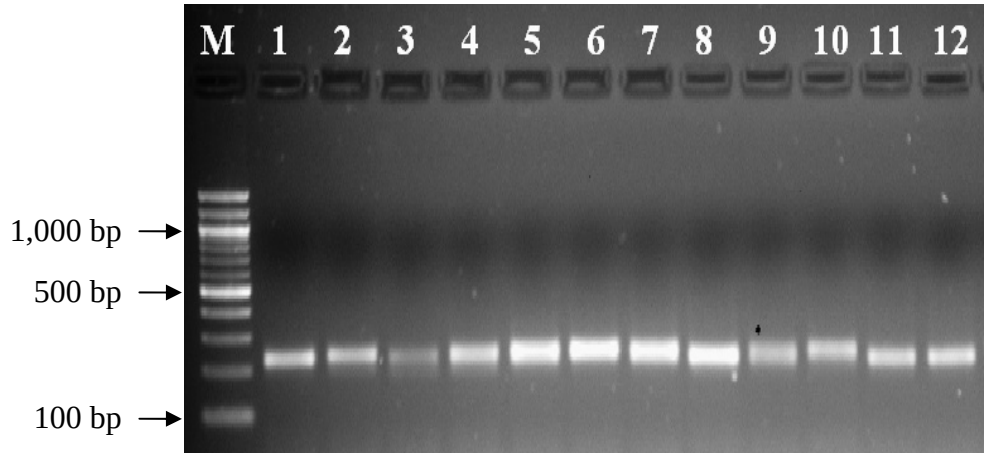

Figure 4.10. Ethidium bromide stained 2.0 % (w/v) agarose gel of specific PCR amplicons generated from the positive cross-species amplification of locus *Jla*106 in three *Johannesteijsmannia* species.

M: 100 bp marker

Lane 1-4: *J. magnifica*

Lane 5-8: *J. perakensis*

Lane 9-12: *J. altifrons*

CHAPTER 5
DISCUSSION

CHAPTER 5

DISCUSSION

5.1 Construction of enriched library

There are many different strategies available for the microsatellite isolation as described in Section 2.6. In this study, enrichment method was adopted to isolate *de novo* microsatellites of *J. lanceolata* because it is fast, efficient and less laboratory equipment required (Zane et al., 2002). A CT-enriched genomic library was constructed to isolate microsatellites of *J. lanceolata* for the first time. By using this enriched library, a high number of microsatellite-containing sequences (79.3 % of non-redundant sequences) were obtained with 33 polymorphic microsatellite loci isolated.

Construction of the CT-enriched library was adapted from Lee et al. (2004b) with a slight modification in the transformation of recombinant plasmids into competent cells. The Lee et al. (2004b) method shows high successful rate in developing new microsatellite markers in various plant species e.g. *Shorea leprosula* (Lee et al., 2004b); *Hopea bilitonensis* (Lee et al., 2004a); *Acer pseudoplatanus* (Pandey et al., 2004); *Cryptomeria japonica* (Tani et al., 2004); *Koompassia malaccensis* (Lee et al., 2006a); *Shorea platyclados* (Ng et al., 2009a); *Gonystylus bancanus* (Ng et al., 2009b); and *Clinostigma* savoryanum (Tani et al., 2009). This method implemented the binding of biotinylated oligonucleotides with streptavidin coated magnetic beads in selective hybridization step (Kandpal et al., 1994) to isolate fragments of DNA containing the microsatellites.

Total genomic DNA were extracted from young leaves of plant for construction of microsatellite-enriched library as they contained more cells per weight and resulted in higher yields of DNA. Besides, young leaves also contain less amounts of polysaccharides or fibres and polyphenolics which made the extraction of DNA less problematic. The genomic DNA of J2 individual was chosen to construct the enriched-library owing to its

higher DNA concentration and relatively less DNA smearing in comparison with J1 individual. Moreover, higher DNA purity was also measured in J2 individual with 1.88 ratio absorbance at 260 nm and 280 nm wavelengths (OD_{260}/OD_{280}). Thus, genomic DNA of J2 individual was chosen to ensure the use of high-quality DNA for library construction.

Basically, there are two options available for DNA fragmentation in the construction of enriched-library. First approach uses restriction enzyme to cut the genomic DNA into average fragments according to the genomic base composition and the enzyme recognition site prior to cassette ligation. Second method involves sonication (Kandpal et al., 1994; Karagyozov et al., 1993) which is less dependent on genomic base composition, but it requires further step in order to obtain blunt-end fragments prior to cassette ligation. This can be done by either filling overhangs with T4 DNA polymerase or discarding them with mung bean nuclease (Zane et al., 2002). In this study, the first approach was implemented because it is easier to perform and less laborious if compared to sonication. Although, restriction enzyme might cause unequal sampling of the genomic regions, it was disregarded in some research and this study as well (Armour et al., 1994; Kandpal et al., 1994; Kijas et al., 1994; Refseth et al., 1997).

Nde II was used to digest the genomic DNA instead of *Sau*3A1 restriction enzyme because it is an isoschizomer to *Sau*3A1 which cleaves at the same recognition site, 5'…↓GATC…3'. *Nde* II is not sensitive to the methylation of its recognition site in contrast to *Sau*3A1 which is sensitive to cytosine methylation (Windhofer et al., 2000). Therefore, more fragments can be obtained when using *Nde* II since it can freely access to its recognition site and cut despite methylation in the site.

DNA fragments ranging from 300 bp to 1,000 bp were recovered after the digestion of genomic DNA. This is because insert DNAs within this range are more prone to be success in cloning afterwards. In addition, primer pairs can be easily designed from the

fragments size within this range. Then, the fragments were ligated to *Sau*3A1 cassettes to facilitate the tagging of unknown DNA fragment sequences. Through cassettes-ligation, the unknown sequences can be enriched for repeats and then amplified by PCR using the cassettes sequences as priming site. Thus, effective ligation of cassettes with DNA fragments is crucial to ensure the quality and enrichment rate of the resulting genomic library (Hamilton et al., 1999).

In this study, selective hybridization were performed using biotinylated $(CT)_{15}$ as probe to fish out the complementary microsatellites-contained sequences. Another method which had used probes cross-linked on a nylon membrane to hybridize with the DNA fragments from the pool was not executed here. This is because the biotinylated probe was fully available for hybridization in liquid medium whereas nylon bound probe had partially cross-linked to the membrane which led to less efficiency in hybridization with target DNA (Zane et al., 2002).

Dinucleotide $(CT)_{15}$ was used as probe in the selective hybridization because according to the *in silico* study of Victoria et al. (2011), the most common dinucleotide microsatellite found in 11 phylogenetically distant plant species were AG/CT and GA/TC. The 11 species were chosen according to the amount of available EST data in NCBI GenBank. They consist of two unicellular green algae (*Chlamydomonas reinhardtii* Dang, *Mesostigma viride* Lauterborn.), three bryophytes s. l. [*Marchantia polymorpha* L., *Physcomitrella patens* and *Syntricha ruralis* (Hedw.) Weber & Mohr], two ferns (*Selaginella* spp. and *Adiantum capillus-veneris* L.), two gymnosperms (*Gnetum gnemon* L. and *Pinus taeda* L.) and two flowering plants with a monocot (*Oryza sativa*) and a dicot (*Arabidopsis thaliana*). In addition, Tang et al. (2010) discovers that the most abundant microsatellites of *Phyllostachys pubescens* searched from NCBI GenBank were dinucleotide microsatellites. Among the dimer motifs, AG/CT or GA/TC was the most

abundant in the database. This similar condition was also observed in rice (*Oryza sativa*) (Victoria et al., 2011). On the contrary, Gupta et al. (2010) revealed AT/TA dinucleotide repeats were the most frequent and followed by AG/CT or GA/TC repeats retrieved from the unigene database of *Solanum lycoperscium* in NCBI. AT-enriched library was not to be used in this study due to the self-complementary nature of AT probes (Lagercrantz et al., 1993; Morgante and Olivieri, 1993; Powell et al., 1996; Stajner et al., 2005). In 2005, several attempts have been done by Stajner et al. (2005) to isolate microsatellites from the AT-enriched library but failed. Hence, dinucleotide $(CT)_{15}$ was used as probe in this study to isolate as many microsatellites as possible.

TempliPhi kit was used to amplify circular because the templiphi reaction discarded the extensive time-consuming step for the bacterial growth in the current plasmid extraction protocol. It utilized bacteriophage Phi29 DNA polymerase and rolling circle amplification for the rapid amplification of circular template DNA. In addition, templiphi required a very small amount of template DNA as low as 1 pg to generate high quality DNA which can be directly added to a sequencing reaction without any purification (Reagin et al., 2003).

5.2 Analysis of sequence data

Clone redundancy is unavoidable in the development of microsatellite-enriched library. The overall redundancy encountered in the library was 15.2 %, which is relatively low compared to other research using the similar enrichment approach. For instance, Lee et al. (2004b) reported 46.8 % clones redundancy in *Shorea leprosula*; Lee et al. (2006a) reported 27.3 % clones redundancy in *Koompassia malaccensis* and Ng et al. (2009a) reported 44.1 % clones redundancy in *Shorea platyclados* for the development of microsatellite markers in respective species.

5.3 Primer design

A total of 111 primer pairs were designed and synthesized from the 353 available microsatellite sequences. This is due to the deficiency of microsatellite flanking regions for primer design. During primer design, perfect microsatellites were prioritized than interrupted and composite microsatellites. Various studies showed that long and uninterrupted or perfect microsatellites were highly mutated than short or interrupted microsatellites (Ellegren, 2000; Schlötterer, 2000; Stajner et al., 2005). The interruption in repeat arrays might serve as anchor point to prevent slippage during DNA replication which led to low polymorphism in the interrupted microsatellites (Chambers and MacAvoy, 2000; Petes et al., 1997). Therefore, probability for long and perfect microsatellites being polymorphic will be much higher than shorter and interrupted microsatellites. Furthermore, scoring of allele size in perfect microsatellites is rather straight forward (e.g. 2 bp differences for perfect dinucleotide microsatellite) in comparison to the interrupted microsatellites.

5.4 Primer screening

Although great efforts were putting in primer design, some microsatellites were not proceeded to polymorphism screening due to the presence of multiple bands, intricated peak patterns, and even unable to amplify in PCR. The multiple bands might be due to segmental duplication events or similarity in priming sites across the genome (Blair et al., 2008). Increasing the annealing temperature up to 10 ºC to avoid non-specific amplification could solve certain occurrence of multiple bands. A possible reason for the PCR failure observed in particular microsatellite loci could be owing to chimeras, instead of real genomic sequences formed during the PCR recovery of DNA prior to cloning. The chimeras were believed to have formed in the later cycles of PCR, when the concentrations of dNTP and primer were low, while the number of complete and incomplete elongated

amplicons increased. Under such circumstances, a significant amount of heteroduplex would be formed (Kanagawa, 2003). These heteroduplexes do not represent a contiguous sequence in *J. lanceolata* genome and therefore primers designed for these chimeras will indeed not amplifiable. Besides, formation of concatamers of multiple fragments during plasmid ligation could also resulted in the PCR failure in the loci (Fischer, 1998).

5.5 Characterization of microsatellite loci

Among the 33 polymorphic microsatellite loci derived from CT-enriched library, majority of them (93.9 %) are composed of AG/CT perfect repeats. Two non-targeted microsatellite loci with TA and TG repeats are surprisingly found in loci *Jla*332b and *Jla*187a respectively. This incidence of isolating the non-targeted microsatellite repeats was solely occurred by chance with low frequency.

All 33 microsatellite loci showed high multi-allelism with average 10 alleles per locus. This value is higher than average number of seven alleles per locus reported by Lee et al. (2006a) with 24 samples screened for polymorphism. Polymorphic information content, PIC is a measure of informativeness related to expected heterozygosity which is derived from allele frequencies. A marker is considered as highly polymorphic if the PIC value is greater than 0.5 (Botstein et al., 1980). Most of the loci (93.9 %) have PIC > 0.5 except for loci *Jla*187a and *Jla*332b. These two loci were the non-targeted TG and TA microsatellites with low number of repeat motifs. According to the study on cultivated peanut or groundnut, *Arachis hypogaea* L. by Cuc et al. (2008), the AG/CT repeat motifs showed higher informativeness than AC/TG repeat motifs. Therefore in order to develop highly polymorphic markers, AG/CT repeat-based microsatellite markers should be prioritized.

Among the 33 microsatellite loci, two loci (*Jla*130b and *Jla*332a) showed departure from Hardy-Weinberg equilibrium from the polymorphism screening of 24 samples. This

might be due to the excessive number of homozygotes in the loci. Besides, Micro-checker analysis suggested the presence of null alleles in both *Jla*332a and *Jla*118 loci which might lead to high number of observed homozygote genotypes. Null allele can be defined as any allele at a microsatellite locus which failed to be amplified in PCR. The potential causes of null alleles are: 1) nucleotide divergence (point mutations or indel) that causes poor primer annealing in either one or both flanking primer regions; 2) short allele dominance where the short allele amplified more efficiently than larger allele; and 3) poor or inconsistent quality of DNA template (Dakin and Avise, 2004). Presence of null alleles can be very problematic in population genetic studies. It underestimated the number of heterozygote in population genetic studies, miscalculated rate of inbreeding in mating system analysis using open-pollinated seeds and underestimated of pollen dispersal distance in paternity analysis (Tani et al., 2004). Therefore, locus with the presence of null alleles will be avoided to be used as marker in the future parentage and population studies.

Besides the presence of null alleles, other factors such as selection pressure, inbreeding and Wahlunds's effect can also contribute to the deviation from Hardy-Weinberg equilibrium (Lee et al., 2006b). However, these three contributing factors cannot be deduced in the study due to the limited number of samples screened.

A number of studies have suggested that the length of microsatellite repeat motifs contributed to the variability of microsatellites (Burstin et al., 2001; Ellegren, 2004; Hüttel et al., 1999; Moretzsohn et al., 2005; Tani et al., 2004; Weber, 1990). Therefore, R program (Team, 2011) was implemented in this study to check the relationship between polymorphism of microsatellite markers and the length of repeat motifs in the corresponding microsatellites. Two scatter plots were made between length of repeat motifs and number of alleles observed per locus (Figure 5.1) and PIC values calculated (Figure 5.2).

The scatter plot in Figure 5.1 showed a positive relationship between length of microsatellite repeat motifs and the number of alleles observed. A statistical relationship can be inferred as the observations were scattered approximately to a line (Le, 2003). As the length of repeat motifs increases, the number of alleles observed per locus increases. Next, Pearson's correlation coefficient, $r = 0.661$ (P-value = 2.816×10^{-5}) was calculated which indicating a positive association. Slope coefficient, β in the linear relationship was subsequently examined to quantify the relationship of length of microsatellite repeat motifs and polymorphic parameters. The estimated slope, $\beta = 0.526$ (P-value = 2.816×10^{-5}) showed a very significant result with 95 % confident intervals between 0.307 and 0.745. Therefore in general, five alleles could be observed per every 10 repeat motifs of a microsatellite locus.

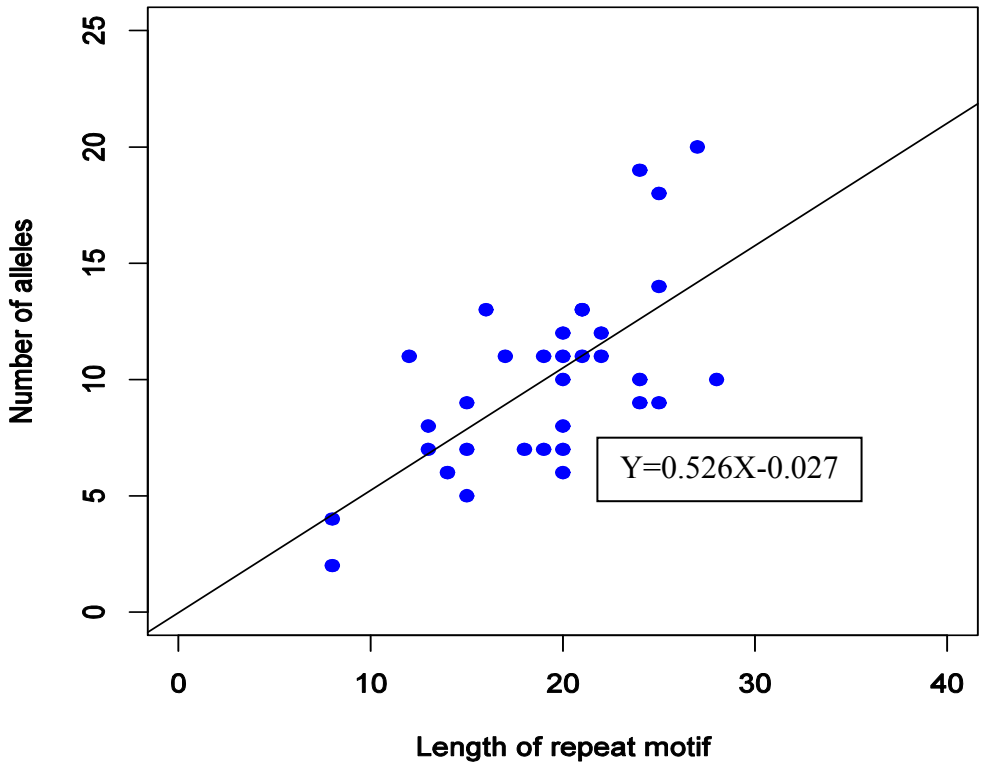

Figure 5.1. Relationships between length of the microsatellite repeat motifs and number of alleles observed per locus.

Figure 5.2 illustrated a positive relationship between length of microsatellite repeat motifs and the calculated PIC values. Most of the observations were scattered around a line except for repeat motifs less than ten with PIC value (< 0.5). The higher PIC values (> 0.6) were between 12 and 28 repeat motifs and clustered more around the line. Pearson's correlation coefficient, r = 0.689 (P-value = 9.147 x 10^{-6}) was obtained suggesting a positive association. The slope coefficient, β of this linear relationship is 0.017 (P-value = 9.147 x 10^{-6}) with 95 % confident intervals between 0.011 and 0.024. Hence, it is possible that up to 0.24 PIC value could be observed in every 10 microsatellite repeat motifs.

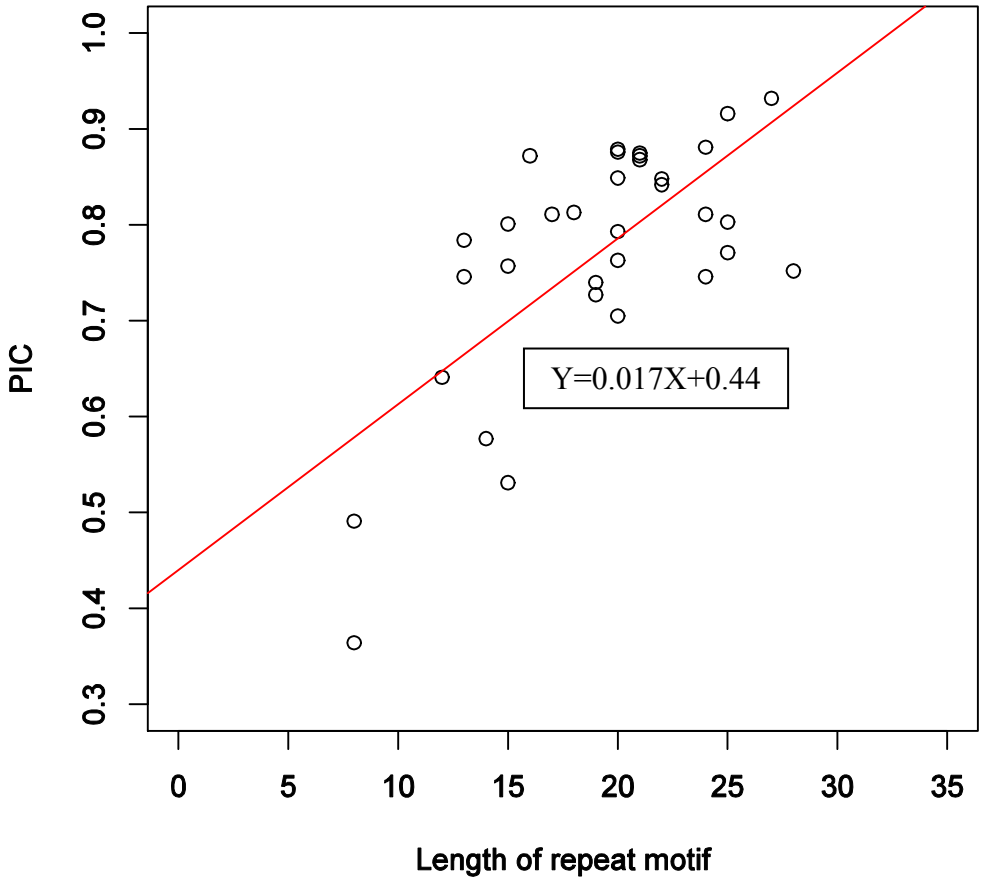

Figure 5.2. Relationships between length of microsatellite repeat motifs and PIC values.

In overall, there is a significant positive relationship between length of microsatellite repeat motifs and level of polymorphism represented by number of alleles observed per locus and PIC values. Similarly, Tani et al. (2004) also detected a significant correlation between the length of repeat units size and the degree of polymorphism derived from PIC and number of alleles per locus in the microsatellite markers developed for sugi (*Cryptomeria japonica*).

5.6 Cross-species amplification

Cross-species amplification or transferability of microsatellites occurred when there is a conserved or homologous microsatellite's flanking region between closely related species. The transferability of microsatellite loci in related species particularly within the same genus has been reported in several studies (Chen et al., 2010; Collevatti et al., 1999; Lee et al., 2006a; Lee et al., 2004a; Ma et al., 2009; Peakall et al., 1998; Sun et al., 2010; Zhai et al., 2010; Zhang and Li, 2010). In this study, high transferability of microsatellite loci developed for *J. lanceolata* was observed in three *Johannesteijsmannia* species (*J. altifrons*, 100 %; *J. magnifica*, 93.9 % and *J. perakensis*, 97.0 %), indicating a high level of genome homology. Thus, it is possible to use these microsatellite loci for future population genetic studies as well as evaluation of the phylogenetic relationships among four *Johannesteijsmannia* species. Prior to this, further evaluation for the polymorphism of each transferable locus in targeted species should be carried out. This is because the transferability of the microsatellite loci in this study was determined based on the PCR amplification in four individuals per species without further polymorphic testing.

5.7 Future plans

A more comprehensive sampling of *J. lanceolata* from its natural populations in Peninsular Malaysia needs to be carried out for the future population genetic studies of the species. However, not all 33 polymorphic microsatellite loci will be used for the studies. This is due to the significant departure from Hardy-Weinberg equilibrium in loci *Jla*130b and *Jla*332a; presence of null alleles in loci *Jla*118 and *Jla*332a; and low PIC (< 0.5) value in loci *Jla*187a and *Jla*332b. Hence, the remaining 28 polymorphic microsatellite loci will be adequate to study the genetic diversity, population genetic structure, patterns of gene flow and mating system of the species. Ultimately, more effective scientific conservation and restoration strategies for the species can be postulated with the genetic information generated. Furthermore, a phylogenetic analysis of *Johannesteijsmannia* species can be executed to investigate species relationships since most of the microsatellite loci are transferable to other *Johannesteijsmannia* species.

CHAPTER 6
CONCLUSION

CHAPTER 6

CONCLUSION

A set of 33 polymorphic microsatellite loci was successfully isolated from a CT-enriched library in *J. lanceolata* via the selective hybridization of biotinylated $(CT)_{15}$ repeats and streptavidin-coated magnetic beads. However, not all the 33 polymorphic microsatellite loci will be used for future studies. This is due to the significant departure from Hardy-Weinberg equilibrium in loci *Jla*130b and *Jla*332a; presence of null alleles in loci *Jla*118 and *Jla*332a; and low PIC (< 0.5) value in loci *Jla*187a and *Jla*332b. The other 28 microsatellite loci exhibited high levels of polymorphism, followed Mendelian codominant inheritance pattern and they also showed the absence of linkage disequilibrium. Hence, they could be used to conduct population genetic studies of *J. lanceolata* by investigating the genetic diversity, population genetic structure, patterns of gene flow and mating system of the species across its distribution range in Peninsular Malaysia. In conjunction with this, more useful information could be generated for the purposes of scientific conservation and restoration strategies for the species. In addition, the majority (> 93.9 %) of microsatellite loci were also transferable to the three species of *Johannesteijsmannia*. Thus, comparable studies and phylogenetic analysis of *Johannesteijsmannia* species could be conducted with these transferable microsatellite loci in the future.

REFERENCES

REFERENCES

Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., . . . Moreno, R.F. (1991). Complementary DNA sequencing: expressed sequence tags and human genome project. *Science, 252*(5013), 1651.

Armour, J.A.L., Neumann, R., Gobert, S., & Jeffreys, A.J. (1994). Isolation of human simple repeat loci by hybridization selection. *Human Molecular Genetics, 3*(4), 599.

Billotte, N., Risterucci, A.M., Barcelos, E., Noyer, J.L., Amblard, P., & Baurens, F.C. (2001). Development, characterisation, and across-taxa utility of oil palm (*Elaeis guineensis* Jacq.) microsatellite markers. *Genome, 44*(3), 413-425.

Blair, M.W., Buendía, H.F., Giraldo, M.C., Métais, I., & Peltier, D. (2008). Characterization of AT-rich microsatellites in common bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics, 118*(1), 91-103.

Blouin, M.S. (2003). DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. *Trends in Ecology & Evolution, 18*(10), 503-511.

Botstein, D., White, R.L., Skolnick, M., & Davis, R.W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of human genetics, 32*(3), 314.

Bowers, J., Boursiquot, J.M., This, P., Chu, K., Johansson, H., & Meredith, C. (1999). Historical genetics: the parentage of Chardonnay, Gamay, and other wine grapes of northeastern France. *Science, 285*(5433), 1562.

Brondani, R.P.V., Brondani, C., Tarchini, R., & Grattapaglia, D. (1998). Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. *Theoretical and Applied Genetics, 97*(5), 816-827.

Burstin, J., Deniot, G., Potier, J., Weinachter, C., Aubert, G., & Barranger, A. (2001). Microsatellite polymorphism in *Pisum sativum*. *Plant Breeding, 120*(4), 311-317.

Butcher, P.A., Decroocq, S., Gray, Y., & Moran, G.F. (2000). Development, inheritance and cross-species amplification of microsatellite markers from *Acacia mangium*. *Theoretical and Applied Genetics, 101*(8), 1282-1290.

Chambers, G.K., & MacAvoy, E.S. (2000). Microsatellites: consensus and controversy. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology, 126*(4), 455-476.

Chan, Y.M. (2009). The Reproductive Biology and Ecology of *Johannesteijsmannia lanceolata* J. Dransf. (Arecaceae). Unpublished master's thesis, University of Malaya, Kuala Lumpur.

Chen, C., Liu, Y.H., Fu, C.X., & Qiu, Y.X. (2010). New microsatellite markers for the rare plant *Cercidiphyllum japonicum* and their utility for *Cercidiphyllum magnificum*. *American Journal of Botany, 97*(9), e82-e84.

Cifarelli, R.A., Gallitelli, M., & Cellini, F. (1995). Random amplified hybridization microsatellites (RAHM): isolation of a new class of microsatellite-containing DNA clones. *Nucleic Acids Research, 23*(18), 3802.

Cipriani, G., Lot, G., Huang, W.G., Marrazzo, M.T., Peterlunger, E., & Testolin, R. (1999). AC/GT and AG/CT microsatellite repeats in peach [*Prunus persica* (L) Batsch]: isolation, characterisation and cross-species amplification in *Prunus*. *Theoretical and Applied Genetics, 99*(1), 65-72.

Collevatti, R.G., Brondani, R.V., & Grattapaglia, D. (1999). Development and characterization of microsatellite markers for genetic analysis of a Brazilian endangered tree species *Caryocar brasiliense*. *Heredity, 83*(6), 748-756.

Cuc, L.M., Mace, E.S., Crouch, J.H., Quang, V.D., Long, T.D., & Varshney, R.K. (2008). Isolation and characterization of novel microsatellite markers and their application for diversity assessment in cultivated groundnut (*Arachis hypogaea*). *BMC Plant Biology, 8*(1), 55.

Dakin, E.E., & Avise, J.C. (2004). Microsatellite null alleles in parentage analysis. *Heredity, 93*(5), 504-509.

Dettman, J.R., & Taylor, J.W. (2004). Mutation and evolution of microsatellite loci in *Neurospora*. *Genetics, 168*(3), 1231.

Dong, Q., Kroiss, L., Oakley, F.D., Wang, B.B., & Brendel, V. (2005). Comparative EST analyses in plant systems. *Methods in enzymology, 395*, 400.

Dransfield, J. (1970). Studies in the Malayan Palms *Eugeissona* and *Johannesteijsmannia*. Ph.D. Thesis, Cambridge University, England.

Dransfield, J. (1972). The genus *Johannesteijsmannia* H.E. Moore, Jr. *Garden's Bulletin Singapore 26*, 63-83.

Dransfield, J., Uhl, Natalie W., Asmussen, Conny B., Baker, William J., Harley, Madeline M., & Lewis, Carl E. (2008). *Genera Palmarum: the evolution and classification of palms*: Kew Pub.

Edwards, K.J., Barker, J.H., Daly, A., Jones, C., & Karp, A. (1996). Microsatellite libraries enriched for several microsatellite sequences in plants. *Biotechniques, 20*(5), 758.

Eisen, J.A. (1999). Mechanistic basis for microsatellite instability. In: Goldstein D. & Schlotterer C., (eds.), *Microsatellites: evolution and applications*. Oxford University Press, Oxford, England. Pp 34–48.

Ellegren, H. (2000). Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genetics, 16*(12), 551-558.

Ellegren, H. (2004). Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics, 5*(6), 435-445.

Eujayl, I., Sorrells, M.E., Baum, M., Wolters, P., & Powell, W. (2002). Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. *Theoretical and Applied Genetics, 104*(2), 399-407.

Fischer, D. (1998). Microsatellite enrichment in organisms with large genomes (*Allium cepa* L.). *Biotechniques, 24*(5), 796-802.

Fisher, P.J., Gardner, R.C., & Richardson, T.E. (1996). Single locus microsatellites isolated using 5 anchored PCR. *Nucleic Acids Research, 24*(21), 4369.

Gaggiotti, O.E., Lange, O., Rassmann, K., & Gliddon, C. (1999). A comparison of two indirect methods for estimating average levels of gene flow using microsatellite data. *Molecular Ecology, 8*(9), 1513-1520.

Garner, T.W.J., Angelone, S., & Pearman, P.B. (2003). Genetic depletion in Swiss populations of *Rana latastei*: conservation implications. *Biological Conservation, 114*(3), 371-376.

Govaerts, R., & Dransfield, J. (2005). *World checklist of palms*: Royal Botanic Gardens, Kew.

Gupta, S., Tripathi, K.P., Roy, S., & Sharma, A. (2010). Analysis of unigene derived microsatellite markers in family Solanaceae. *Bioinformation, 5*(3), 113.

Gur-Arie, R., Cohen, C.J., Eitan, Y., Shelef, L., Hallerman, E.M., & Kashi, Y. (2000). Simple sequence repeats in *Escherichia coli*: abundance, distribution, composition, and polymorphism. *Genome Research, 10*(1), 62.

Hamilton, M.B., Pincus, E.L., Di Fiore, A., & Fleischer, R.C. (1999). Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. *Biotechniques, 27*(3), 500-507.

Hüttel, B., Winter, P., Weising, K., Choumane, W., Weigand, F., & Kahl, G. (1999). Sequence-tagged microsatellite site markers for chickpea (*Cicer arietinum* L.). *Genome, 42*(2), 210-217.

Isagi, Y., & Suhandono, S. (1997). PCR primers amplifying microsatellite loci of *Quercus myrsinifolia* Blume and their conservation between oak species. *Molecular Ecology, 6*(9), 897-899.

Johnson, D.V. (1996). *Palms: their conservation and sustained utilization : status survey and conservation action plan*: IUCN, Gland, Switzerland and Cambridge, UK: Appendix 2.

Jongeneel, C.V. (2000). Searching the expressed sequence tag (EST) databases: panning for genes. *Briefings in Bioinformatics, 1*(1), 76.

Kalinowski, S.T., Taper, M.L., & Marshall, T.C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology, 16*(5), 1099-1106.

Kanagawa, Takahiro. (2003). Bias and artifacts in multitemplate polymerase chain reactions (PCR). *Journal of Bioscience and Bioengineering, 96*(4), 317-323.

Kandpal, R.P., Kandpal, G., & Weissman, S.M. (1994). Construction of libraries enriched for sequence repeats and jumping clones, and hybridization selection for region-specific markers. *Proceedings of the National Academy of Sciences of the United States of America, 91*(1), 88.

Karagyozov, L., Kalcheva, I.D., & Chapman, V.M. (1993). Construction of random small-insert genomic libraries highly enriched for simple sequence repeats. *Nucleic Acids Research, 21*(16), 3911.

Kiew, R. (1991). Palm utilization and conservation in Peninsular Malaysia. In: Johnson D. (ed.), *Palms For Human Needs In Asia – Palm utilization and conservation in India, Indonesia, Malaysia and the Philippines*. A.A. Balkema, Rotterdam. Pp 75-130.

Kijas, J.M.H., Fowler, J.C.S., Garbett, C.A., & Thomas, M.R. (1994). Enrichment of microsatellites from the citrus genome using biotinylated oligonucleotide sequences bound to streptavidin-coated magnetic particles. *Biotechniques, 16*(4), 656-662.

Krutovsky, K.V., St. Clair, J.B., Saich, R., Hipkins, V.D., & Neale, D.B. (2009). Estimation of population structure in coastal Douglas-fir [*Pseudotsuga menziesii* (Mirb.) Franco var. *menziesii*] using allozyme and microsatellite markers. *Tree Genetics & Genomes, 5*(4), 641-658.

Kumar, S.V., Tan, S.G., Quah, S.C., Yusoff, K. (2002a). Isolation of microsatellite markers in mungbean, *Vigna radiata*. *Molecular Ecology Notes,* 2:96-98.

Kumar, S.V., Tan, S.G., Quah, S.C., Yusoff, K. (2002b). Isolation and characterization of seven tetranucleotide microsatellite loci in mungbean, *Vigna radiata*. *Molecular Ecology Notes,* 2:293-295.

Lagercrantz, U., Ellegren, H., & Andersson, L. (1993). The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. *Nucleic Acids Research, 21*(5), 1111.

Le, C.T. (2003). *Introductory biostatistics*: Wiley-Interscience.

Lee, C. T., Lee, L., Faridah, Q. Z., Siraj, S. S., Ng, K. K. S., Norlia, B., & Mat-Isa, M. N. (2006a). Isolation and characterization of microsatellite markers in *Koompassia malaccensis* (Leguminosae), an important tropical timber species. *Molecular Ecology Notes, 6*(4), 1198-1201.

Lee, S. L., Tani, N., Ng, K. K. S., & Tsumura, Y. (2004a). Characterization of 15 polymorphic microsatellite loci in an endangered tropical tree *Hopea bilitonensis* (Dipterocarpaceae) in Peninsular Malaysia. *Molecular Ecology Notes, 4*(2), 147-149.

Lee, S. L., Tani, N., Ng, K. K. S., & Tsumura, Y. (2004b). Isolation and characterization of 20 microsatellite loci for an important tropical tree *Shorea leprosula* (Dipterocarpaceae) and their applicability to *S. parvifolia*. *Molecular Ecology Notes, 4*(2), 222-225.

Lee, Yuan-Jr, Hwang, Shih-Ying, Ho, Kuo-Chieh, & Lin, Tsan-Piao. (2006). Source populations of *Quercus glauca* in the last glacial age in Taiwan revealed by nuclear microsatellite markers. *Journal of Heredity, 97*(3), 261-269.

Li, W., & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics, 22*(13), 1658.

Litt, M., & Luty, J.A. (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of human genetics, 44*(3), 397.

Liu, Z.J., & Cordes, J.F. (2004). DNA marker technologies and their applications in aquaculture genetics. *Aquaculture, 238*(1-4), 1-37.

Look, S.L. (2007). Population genetics and phylogeny of the malesian palm genus *Johannesteijsmannia* H.E.Moore (Palmae). Umpublished PhD's thesis, National University of Singapore, Singapore.

Lunt, D.H., Hutchinson, W.F., & Carvalho, G.R. (1999). An efficient method for PCR-based isolation of microsatellite arrays (PIMA). *Molecular Ecology, 8*(5), 891-894.

Ma, H., Chen, S., Liao, X., Xu, T., & Ge, J. (2009). Isolation and characterization of polymorphic microsatellite loci from a dinucleotide-enriched genomic library of obscure puffer (*Takifugu obscurus*) and cross-species amplification. *Conservation Genetics, 10*(4), 955-957.

Moretzsohn, M.C., Leoi, L., Proite, K., Guimaraes, P.M., Leal-Bertioli, S.C.M., Gimenes, M.A., . . . Bertioli, D.J. (2005). A microsatellite-based, gene-rich linkage map for the AA genome of *Arachis* (Fabaceae). *Theoretical and Applied Genetics, 111*(6), 1060-1071.

Morgante, M., & Olivieri, A.M. (1993). PCR amplified microsatellites as markers in plant genetics. *The Plant Journal, 3*(1), 175-182.

Morgante, M., Hanafey, M., & Powell, W. (2002). Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nature Genetics, 30*(2), 194-200.

Murray, M.G., & Thompson, W.F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research, 8*(19), 4321.

Ng, C.H., Lee, S.L., Ng, K.K.S., Muhammad, N., & Lee, C.T. (2009). Development of microsatellite markers for *Shorea platyclados* (Dipterocarpaceae). *Conservation Genetics Resources, 1*(1), 317-319.

Ng, K.K.S. (2005). Spatial structure and impact of logging on genetic diversity of selected tropical tree species. Unpublished PhD's thesis, University of Malaya, Kuala Lumpur.

Ng, K.K.S., Lee, S.L., Ng, C.H., Tnah, L.H., Lee, C.T., & Tani, N. (2009). Microsatellite markers of *Gonystylus bancanus* (Thymelaeaceae) for population genetic studies and DNA fingerprinting. *Conservation Genetics Resources, 1*(1), 153-157.

Oliveira, E.J., Pádua, J.G., Zucchi, M.I., Vencovsky, R., & Vieira, M.L.C. (2006). Origin, evolution and genome distribution of microsatellites. *Genetics and Molecular Biology, 29*(2), 294-307.

Pandey, M., Gailing, O., Fischer, D., Hattemer, H. H., & Finkeldey, R. (2004). Characterization of microsatellite markers in sycamore (*Acer pseudoplatanus* L.). *Molecular Ecology Notes, 4*(2), 253-255.

Parker, P.G., Snow, A.A., Schug, M.D., Booton, G.C., & Fuerst, P.A. (1998). What molecules can tell us about populations: choosing andusing a molecular marker. *Ecology, 79*(2), 361-382.

Peakall, R., Gilmore, S., Keys, W., Morgante, M., & Rafalski, A. (1998). Cross-species amplification of soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: implications for the transferability of SSRs in plants. *Molecular biology and evolution, 15*(10), 1275.

Petes, T.D., Greenwell, P.W., & Dominska, M. (1997). Stabilization of microsatellite sequences by variant repeats in the yeast *Saccharomyces cerevisiae*. *Genetics, 146*(2), 491.

Powell, W., Machray, G.C., & Provan, J. (1996). Polymorphism revealed by simple sequence repeats. *Trends in plant science, 1*(7), 215-222.

Primmer, C.R., & Merilä, J. (2002). A low rate of cross-species microsatellite amplification success in Ranid frogs. *Conservation Genetics, 3*(4), 445-449.

Rafalski, J.A. (2002). Novel genetic mapping tools in plants: SNPs and LD-based approaches. *Plant Science, 162*(3), 329-333.

Rakoczy-Trojanowska, M., & Bolibok, H. (2004). Characteristics and a comparison of three classes of microsatellite-based markers and their application in plants. *Cellular and molecular biology letters, 9*(2), 221-238.

Reagin, M.J., Giesler, T.L., Merla, A.L., Resetar-Gerke, J.M., Kapolka, K.M., & Mamone, J.A. (2003). TempliPhi: a sequencing template preparation procedure that eliminates overnight cultures and DNA purification. *Journal of Biomolecular Techniques: JBT, 14*(2), 143.

Refseth, U.H., Fangan, B.M., & Jakobsen, K.S. (1997). Hybridization capture of microsatellites directly from genomic DNA. *Electrophoresis, 18*(9), 1519-1523.

Richardson, T., Cato, S., Ramser, J., Kahl, G., & Weising, K. (1995). Hybridization of microsatellites to RAPD: a new source of polymorphic markers. *Nucleic Acids Research, 23*(18), 3798.

Roa, A.C., Chavarriaga-Aguirre, P., Duque, M.C., Maya, M.M., Bonierbale, M.W., Iglesias, C., & Tohme, J. (2000). Cross-species amplification of cassava (*Manihot esculenta*)(Euphorbiaceae) microsatellites: allelic polymorphism and degree of relationship. *American Journal of Botany, 87*(11), 1647-1655.

Rousset, F. (2008). Genepop'007: a complete re-implementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources, 8*(1), 103-106.

Rozainah, M.Z. & Sinniah, U.R. (2005). Population structure and spatial distribution of umbrella leaf palm at Angsi Forest Reserve, Malaysia. *Malaysian Forester, 68*(2), 98-104.

Rozainah, M.Z., & Sinniah, U.R. (2006). Flowering event of *Johannesteijsmannia lanceolata*: An understorey palm in the Angsi Forest Reserved, Malaysia. *Malaysian Journal of Science, 25*(1), 65-68.

Rudd, S. (2003). Expressed sequence tags: alternative or complement to whole genome sequences? *Trends in plant science, 8*(7), 321-329.

Saw, L.G. (1998). Palms. In Soepadmo E. (ed.), *The Encyclopedia of Malaysia Vol. 2-Plants.* Archipelago Press, Singapore. Pp 54-55.

Schlötterer, C. (2000). Evolutionary dynamics of microsatellite DNA. *Chromosoma, 109*(6), 365-371.

Sia, E.A., Butler, C.A., Dominska, M., Greenwell, P., Fox, T.D., & Petes, T.D. (2000). Analysis of microsatellite mutations in the mitochondrial DNA of *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America, 97*(1), 250.

Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics, 139*(1), 457.

Spooner, D.M., Treuren, R., & Vicente, M.C. (2005). *Molecular markers for genebank management*: IPGRI.

Stajner, N., Jakse, J., Kozjak, P., & Javornik, B. (2005). The isolation and characterisation of microsatellites in hop (*Humulus lupulus* L.). *Plant Science, 168*(1), 213-221.

Sun, Y., Fu, C.X., & Qiu, Y.X. (2010). Development, characterization, and transferability of microsatellite markers for *Kirengeshoma palmata* (Hydrangeaceae). *American Journal of Botany, 97*(6), e48-e51.

Tang, D.Q., Lu, J.J., Fang, W., Zhang, S., & Zhou, M.B. (2010). Development, characterization and utilization of GenBank microsatellite markers in *Phyllostachys pubescens* and related species. *Molecular Breeding, 25*(2), 299-311.

Tani, N., Ohtani, M., & Yoshimaru, H. (2009). Development and diversity of microsatellite markers for endangered species *Clinostigma savoryanum*. *Conservation Genetics, 10*(5), 1521-1523.

Tani, N., Takahashi, T., Ujino-Ihara, T., Iwata, H., Yishimura, K., & Tsumura, Y. (2004). Development and characteristics of microsatellite markers for sugi (*Cryptomeria japonica* D. Don) derived from microsatellite-enriched libraries. *Annals of Forest Science, 61*(6), 569-575.

Tautz, D. (1989). Hypervariabflity of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research, 17*(16), 6463.

Tautz, D., & Schlotterer, C. (1994). Simple sequences. *Current Opinion in Genetics & Development, 4*(6), 832-837.

Team, R. (2011). R: A language and environment for statistical computing. *R Foundation for Statistical Computing Vienna Austria ISBN, 3*(10).

Thiel, T., Michalek, W., Varshney, R., & Graner, A. (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics, 106*(3), 411-422.

Uhl, N.W. & Dransfield, J. (1987). *Genera Palmarum*. Allen Press, Lawrence.

Van Oosterhout, C., Hutchinson, W.F., WILLS, D.P.M., & Shipley, P. (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes, 4*(3), 535-538.

Victoria, Filipe, da Maia, Luciano, & de Oliveira, Antonio. (2011). In silico comparative analysis of SSR markers in plants. *BMC Plant Biology, 11*(1), 15.

Vienne, D. (2003). *Molecular markers in plant genetics and biotechnology*: Science Publishers.

Vignal, A., Milan, D., SanCristobal, M., & Eggen, A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution, 34*(3), 275-306.

Walter, K.S., & Gillett, H.J. (1998). *1997 IUCN red list of threatened plants*: IUCN-The World Conservation Union, Gland, Switzerland and Cambridge, UK.

Weber, J.L. (1990). Informativeness of human (dC-dA)n·(dG-dT)n polymorphisms. *Genomics, 7*(4), 524-530.

Weir, B.S. (1996). *Genetic data analysis 2*: Sinauer Associates.

White, G., & Powell, W. (1997). Isolation and characterization of microsatellite loci in *Swietenia humilis* (Meliaceae): an endangered tropical hardwood species. *Molecular Ecology, 6*(9), 851-860.

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A., & Tingey, S.V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research, 18*(22), 6531.

Windhofer, F., Catcheside, D.E.A., & Kempken, F. (2000). Methylation of the foreign transposon Restless in vegetative mycelia of *Neurospora crassa*. *Current Genetics, 37*(3), 194-199.

Wirgin, I.I., & Waldman, J.R. (1994). What DNA can do for you. *Fisheries, 19*(7), 16-27.

Wu, K., Jones, R., Danneberger, L., & Scolnik, P.A. (1994). Detection of microsatellite polymorphisms without cloning. *Nucleic Acids Research, 22*(15), 3257.

Zane, L., Bargelloni, L., & Patarnello, T. (2002). Strategies for microsatellite isolation: a review. *Molecular Ecology, 11*(1), 1-16.

Zhai, S.N., Yan, X.L., Nakamura, K., Mishima, M., & Qiu, Y.X. (2010). Isolation of compound microsatellite markers for the endangered plant *Neolitsea sericea* (Lauraceae). *American Journal of Botany, 97*(12), e139-e141.

Zhang, L., & Li, Q.M. (2010). Isolation and characterization of microsatellite markers in an endangered species *Dracaena cambodiana* (Liliaceae). *American Journal of Botany, 97*(10), e91-e93.

Zucchi, M.I., Brondani, R.P.V., Pinheiro, J.B., Brondani, C., & Vencovsky, R. (2002). Transferability of microsatellite markers from *Eucalyptus* spp. to *Eugenia dysenterica* (Myrtaceae family). *Molecular Ecology Notes, 2*(4), 512-513.

APPENDICES

APPENDIX A: DNA sequences derived from microsatellite-containing clones of *Johannesteijsmannia lanceolata* with respective clone name and followed by the NCBI GenBank Accession number. The microsatellites are in bold.

*Jla*002 HQ613842

```
   1 cttgtgatta ggagtggtat gataaccatg atgtaccgag ccagggaata tattgggtgg
  61 acggccatgt tttcaaataa gatggataaa aagaacactt cgttcttatt aggctgaaca
 121 attctatctc ttatatttat gcttagtgag caaggaaatc aggttctgtg gaagtgcaga
 181 gcgataatta aagaagatgc atctctgtgg tgaaggtgtg gtgcaaggga gtatgtgtta
 241 tgattatata gatgtatgta catacattgt tgtatgaggg gtacggtggt ggatgcatct
 301 ggttgtgtgt gtacacgtgc actcttgcgc acaggcatgt gcaaacgtgc atgcacgaga
 361 gagagagaga gagagagaga gagagagacc tgaagcacac tgacgtccaa gtagatgaga
 421 tc
```

*Jla*046 HQ613843

```
   1 ctttttttt agtcccatat cagttattca ctggataaac cttggatgct tatataggac
  61 caagaaatct aaatagcgat tttggatgag gtcttgagtt gttacattgc ctatgtaaaa
 121 tgttaactaa attcataatg aagtcgagaa aatataaaat taattataca ttttcatctt
 181 accattactt tatgtactct ctctctctct ctctctctct ctctctctct ctctctctct
 241 ctctcatgtt taggagtggc tattcaccac ctgctacctc atatccacct catccctcaa
 301 gggccaaagc caactggaat cgaaccctca ccctctgaag agatgatacc atccaagcaa
 361 gcaagtattt gatc
```

*Jla*048 HQ613844

```
   1 ctcttccctt ctttgcacga tttcgagggc acagagcgtg tgaggagaga gagagagaga
  61 gagagagaga gagagccggc atgtagggtt gggtccagcc ctcgtacacc cgaaccggta
 121 cccagacaag taaaccctac caaccctccc atgcttgggt atggagtttt ggattagatg
 181 agaaagtcaa attgatc
```

*Jla*106 HQ613845

```
   1 tccaacgcat tgttctagtg aaaagggttg aagctgctat ttcacataga ttagtgagac
  61 ttggtcatta tatactggca gccttaacgg cttaagccac gagtaggaca tgcatatctg
 121 gatgaagttg acaagttttc ttggtaataa atgatactca catgcaatca ccagcacagg
 181 agaaaatatc aaattttaca taggcttagc cgtaagcccg accaaaagtg aatagaatct
 241 aagtccgagg cccggcccat gattagcttt aaccatggtt tgctactcga atgttctccc
 301 ctcagctatt ctaatagttg ttccatatct ctctctctct ctctctctct ctctctctct
 361 ctctctctca taacacatca tttgaaatct atgcacacat tttgtttgtc ccaacaatac
 421 caaggataaa aaatggaaaa agaaaactca ttctcctaag aagtcaagat gcttcaagtc
 481 tgagggaggc aaacatatag atagataagg gagagatgga tc
```

*Jla*112 HQ613846

```
   1 tcttagttag gcacaacgca ggcttgcaaa agagtccttc cagcctcatt ccaatgaaag
  61 caaatgtact ttccctgagc tctactttgt tactgttgtt cataacatgc atggtgaaag
 121 cctctctctc tctctctctc tctctctctc tctctctctc tctaagttgt gtggtggaga
 181 agttggtgcc atgtggatga gctttaatat ttgggatacc tgcctccacc gtcccttatt
 241 ggctaagaat tcatgatgtg gtgataagat tggcagtgat gatgaccaag tgttggatac
 301 tttaaatcat tgttccttct tatgtgaagg ctactgtttg gagagtcgca ctcattgcat
 361 cgtctggttt ggcagtggac caagccagct ggttagctac gatc
```

*Jla*117 HQ613847

```
   1 agagaaagat tggattttgg taaggaaaac agaggttttg gcaatggaga gaacgagaac
  61 aggggagttt tggtttgtgc gttaacggag aaagagggtg gaagaaagag gagacaaaga
 121 ggccaagcag agagagagag agagagagag agagagtaat tggagagata tatgagaggc
 181 ggtgaccttt tccatggtgt ttgagttgga tggaagaagg tttggcccgt gggtgtgcac
 241 tgggatttga cgcccgggga ggggggagga acgcttgaaa ttatttggcc gcatgtgaac
 301 taaagttttg gctatctgat c
```

*Jla*118 HQ613848

```
   1 tcgccccacg gtaaccttcg tccaactctc tctcccttcc ccatcctttc cgctctctct
  61 ccttcctttc tttctctctt gcaggtcctt cccggggtcc ccgtcccatg aggaccggcc
 121 tctcctcgaa gcactccaag aagaaccccg atgaggtgcc ctctctctct ctctctctct
 181 ctctctctct ctctctctct ctcttcaagt acttcgaact ccgaattctt tctaacctgc
 241 aattactgga ttcatcctta accaggagcg gagcaaaaag agaaagatga gatc
```

*Jla*124 HQ613849

```
   1 tatctgctat ttcaggtgtt tgtagaatag tgaatttgag atgtgctgtt attatgttag
  61 attattatga ggagggctgc aattcagggg tttcttaact tgtatagaaa ttttttcgtc
 121 cagaagtggc ctttgaatta tgtcctgtaa acaacaacca tcaatgtaga aattttgctg
 181 aaatcactat ttctctattg atgttaaatg tgaacctctc tctctctctc tctctctctc
 241 tctctctctc tctcgcccgc tctcgtgcgc gcgtgcgcgc acatttcttc tccaagattc
 301 aacgtatatt tataaactaa ttgttgtagc catggaggcc tatgaaatag attacaaata
 361 agaaattgtg ttctacttaa aattcagata tttcctgtta taacataaga gattttgctc
 421 tacatagtct ggaaatgttc aaattattta ggattttgat gtttgcattt aaatatctca
 481 agatc
```

*Jla*130 HQ613850

```
   1 tggcagtatc gatggaagca aacatcaagg tcccatgttc aacaaccatt tcagttattg
  61 tgtttgctct cttgactcgt ttgagttatg aatgtgtgtg tgtgtgtgtg tatttttttc
 121 tccaatagaa atgatgggga cataaagtag atagtatgaa gattaaagat tttccaaatc
 181 atggtagtgg aggctgctat tgtaaacatt cattgacagt aaattactca agtccaactt
 241 ggcctatgct cacagtcctc tctctctctc tctctctctc tctctctctc tctctctctc
 301 tctcttaagg cgtggtccta agcctctata ctacagttgc tttacttaat gggtccctag
 361 tgtacttcta tgtaccatac tgaagggttg gagggtgaag aaggtgacac cttccattgc
 421 tctccaagta cgggccagat gatttgaggg cctttgccga gatc
```

*Jla*131 HQ613851

```
   1 acatttaacc cataaatcca tcactttgta caaacatatt gcaatttagc tcgttgggac
  61 aataaatcat gaacaattat tctctctctc tctctctctc tctctctctc tctctctctc
 121 tctctctctc tctctctgcg tgtgcatgtg catgtctgtg tgtctgtgtg cgcctttatt
 181 tgcatcaacc actttacata tttcatatgt cttcaactct ctcctttgga gaaaccttag
 241 ggtatagatc
```

*Jla*160 HQ613852

```
   1 gtaaatagta ctttctatat ctgcctcccc cctcccctcc tctctgtctc cccatttctc
  61 tctctctctc tctctctctc tctctctggc ctttgattgc gtgctgtttg gggagttcgg
 121 agagatgaag gttagaaatt atttgattcc ttcttgttta tcaaatgaga tgatattcta
 181 atggcatggc ttctttatgg agtctctcct catcttcatc acttttcagg acacatatct
 241 taaaaaggaa tcgtctttac tgtctagctt gtatttttga tc
```

*Jla*162 HQ613853

```
   1 tctaaaacat gaaaaatctt ctctctcaac atctcttttt tgtccatgtc ttcagatata
  61 gctcttacct ttcatccagc tctaatttag gtgctccacg cccccacctt aagccctgag
 121 tgcctaacct atcaactcac gacgtgtttc attccatgta caaaaggcct cagcctagct
 181 ggttacgctc tcagtacaag ccttacgacc ttgaacgtta ggcgttaacc ttgtcctatg
 241 catacaccac acccacagcc ccactgattg atttcctgat atcatgtaga acgagaacat
 301 atgtcgactg tattatatta cacatacgaa gggagatatc agaatctctc tctctctctc
 361 tctctctctc tctctctctc tctctctctc tctctgtcgc tttcagtgac ggcacggact
 421 gtacacacac agaagattac agcgtacgaa tacggccacc ttatcgccca tctcttttc
 481 tctatatttg ttatatattc tctcgctatc tatcacaatt caccaccagg gatc
```

*Jla*168 HQ613854

```
   1 tcacaagcag ctccacttct tgtactgatg cctcccctca ttggtccccc ttctccttct
  61 ctctctcaaa agagagtgcc cacttgagtc aggacttgaa ttccccatta gagagagaga
 121 gagagagaga gagagagaga gagagagaga gagagagaga gttctgtgga aataatgtat
 181 aagtgtggat gttatatata tggcatgatg gatggaagaa agttcaaggc tagttcctga
 241 ttagacttct agtataaggg tctgatgctt gttggaccat ttgattggcc aatgccatgt
 301 agccaattgc atgtgctaga gtgacatgtg tgtcacatga tgaacagcac attgatcgat
 361 ctagggctat gacaatgtta acttttctga gttgggaaaa tcaccataag acttttgca
 421 gtcctagtag aacctttacc cactctctct accttcttct tcttcttctt cttcttcttc
 481 ttcttcttgt tttctctctc tctctctctc tctctctctc tctctctctc tctctctctc
 541 tc
```

*Jla*174 HQ613855

```
   1 cactctgctg tgcaagctta cttgcttgga gcagaggtaa tgcaaaatca acccctctct
  61 ctctctctct ctctctctct ctctctctct ctgttgcaca ctgcataaaa tagggccata
 121 tctgcaataa ctctcatcat taactatagt aatcaataaa ccacagaaaa tgtataaaat
 181 gttgtaatct gcataacaca gggagcatag tagtgatcga tctccccggc ggctttatac
 241 tgtggcccat tatcagtgat gacctgcacg acgttctcct ctcctatcta atcaatcacc
 301 tcctccatca atctgaggat gtaggagcta tcgtgcacct tatcagaagc atcaatcgac
 361 ttctggaaga aagtcttcgt atcacaatat gtcagaaagt tgatgatgct ccgtctagta
 421 ggaccggtcc aaccgtcaca catcactgtc agtccgtatg taggccactt gctcttgtat
 481 gaggcaatcc acttctcaag ttcttcttta ttactgtcaa gaagctcacc gtagatggcc
 541 ctcaaacccg gaggatc
```

*Jla*186 HQ613856

```
   1 acgagcatct caaaaggaat atcttaacta ccttgattgg tggatatgga aagagtgcaa
  61 ctatttatcc tcgaatcgtg taattctaac catacatatt cgcatcttct cgtttataaa
 121 taccaaacac ggacattctc aagatatgat gggaaattta aataccatac tctcatacct
 181 ctctctctct ctctctctct ctctctctct cttcatctc ttctgttttc tcttggaact
 241 ctgcttgact taagcattaa agggccccca ggacatattc caatgagccg ttgatc
```

*Jla*187 HQ613857

```
   1 ggctccaaag gcttggtctc ggcgctagag gttccctagc gccaagccgg tctcggtctc
  61 ggcacgcgcc taccccgggc ctagtctcag cttactctag aagctaactt tgatgcagag
 121 ccaggcgaat tgcaaaattg gaagcacata aacagaaaaa taatcaccga actcttcgag
 181 tgtggcatca aatcccgtga tcgatcccac tcccaattaa ttaatgatta attgtgttgt
 241 tgtgtgtgtg tgtgcttgtg tgcatgcgtg tgcgtgcgac acaaagagag agagagagag
 301 agagagagag agagagagag agcaaacaaa atagccaatt tatcagataa actgctaaga
 361 ataaaaaggc aatggtaggt agccttcaac agagtcaacc ggagacatta gtaccaaata
 421 tatgatattt tgtgaacaaa aattagtata tgacccgaac cagatggtac gaagcacacc
 481 gtaccggttt ggtactggta tccggtatag gaggtgtatc gacactcaat atgccgaaaa
 541 ggccccatac catactatac caacacagta ctagtgtggc cccagtacgg ggcctggtat
 601 tgagacagcg aaccttattc ttaacattcg aaagctgtct tttcttcata caatatgaat
 661 aaagggacca tgagcc
```

*Jla*192 HQ613858

```
   1 cgtacaacta tagcatcata attatttata ttttttacat gttttggtat caggttttat
  61 gattctgtgt tctgtgtcct ctctctctct ctctctctct ctctctctct ctctctctct
 121 ccccccctct cttttctttg ttgggagaaa tagcgtcctc tggttcaaaa tgttgctagc
 181 ctattcctcg tgttgcttct tttgtgattt gacatttgta tgttgcaatg gttggaagtg
 241 ctcaaaattt ttgacttctt atactaaaac ttctttaatt catcaatcat tgtggctaaa
 301 tgtgattctt gggttacttt tttattctat ttaaactgag caatggttaa tactgtcatt
 361 tcaatttcat atgtattggt tgtatgaatt tattaccttt agagatacac gttgcc
```

*Jla*196 HQ613859

```
   1 atataagaaa ggaaattttc acagaatcta tcagatatgt gaatgacagt aacttaaaat
  61 gtacccaaat cacaactcaa aaagctaggg gcatttgcac attcacatat aaaaattata
 121 tatcttgaag ccttgtgtca gtgaattagt actctctctc tctctctctc tctctctctc
 181 tctctctctc tctctctctc tctctcgtgc gcgcgagctt gccattgtgt atccgtgcac
 241 tcgcttgtaa aaacggtttg acataccata tcttctcgct tgtaaaaacg gtttgacata
 301 ccatagcttg attagcagct ttgcttatgt tagaaattca tttgtaattc tagaagcttt
 361 cttgatcgat ctatggacct gcgattcctc aaatcttctt gcatggaggt attgtgtgag
 421 taggcatact ttccgtaatc agaagtcttc taagttaatt gttctctgtc atttcttgaa
 481 gtaggacctc aatattttga attccattta tgctttgata tttcaataac cattgctaat
 541 ttttatatgt tcatcatcat tgagctttcc catgtctttg gtctaccatc cagtgttggg
 601 ttggccaagc tctctttggt agtttttcct gtcagtctct ctctctctct ctctctctca
 661 ctctctctct
```

*Jla*198 HQ613860

```
   1 ggagtcaaca tgaaattatc atcaaaatca agaatcttat ggcccatcca tccttttttat
  61 tcgatttttc tgaagaatgc agcaattcgt ttgaagatgt ttcaaaattt ttcagactca
 121 aaacaataaa aatgaaattc tactaaagtg gcaattcagc aagaataaga gagagagaga
 181 gagagagaga gagagagaga gagagagaga gtatctcagt cactaaacac tctaaatact
 241 gtgatttaac atgaatcttt tcattccttg cacgtattat tcatccgaaa gagtgttctc
 301 tttccggaca tacatcttga ttctggtgag aacaaaataa tcgaacccta ccgccgccgc
 361 attcaaaatt cccggcaagc ggccggaaaa tcc
```

*Jla*210 HQ613861

```
   1 tacccataat aaaattatat aattacaaaa catgtccctt tatatcataa cgcagcgaga
  61 aaagaattgt gaagacgaaa ggagaagaac actcccagga gccctaagga acctttgtca
 121 gccttggttt ctctgttgtc acgtgataaa gatattattt cttttcttct agagattttg
 181 atttctttta taggagaatt ttgccgtctc catcgccctc acattgttcc ccccgccccc
 241 ccccccctct ttttattgtc ctctacttcc cacaacaatc ggaaaatatg cccacttcct
 301 ctcttcccat ctctttattc tctctcatgc caacattctc cctctctctc tctctctctc
 361 tctctctctc tctctctctc tctcttggca tgttgcggtc cactgtagac cacgacattg
 421 cgatcgatca aaatccctat gttgaccgat ttta
```

*Jla*236 HQ613862

```
   1 cctttgtgaa agtagactcc aaggccatga tttcccttca gtccttgtct ctcatgcctt
  61 ctcaggcctg attccactgc actagattcc tctctctctc tctctctctc tctctctctc
 121 tctctctctg tatacccctt tcgttttctc atgcgcttcc atcaattttt tttttccttt
 181 tttcctctga gactcttctc tttttctccc cctatcttgg tttgatttct atctgtagga
 241 ctggttttga aagcaagtta gaagaggaaa aggagtggga atgatggagg gggacacctt
 301 ctcaggcgtt ggcaatggaa cccaagcgga caacaaggtc ctccaggtct tcaagaagag
 361 ctttgtccaa gtgcagagca tcctggatc
```

*Jla*238 HQ613863

```
   1 cacaactaca attaaacaaa tattgccaat aagaatagta aatgaggtgt tgggtcacta
  61 tgtgattatc tagagtggac aaggattgag gttaattgga ctccttttca caagattttt
 121 agcaggtagg atgttgtttg tgtgtcccat atgttgattg aaaagtatcc aagcagctac
 181 accctctcta tgctttgagg tgcaagaact cccatctctc tctctctctc tctctctctc
 241 tctctctctc tctctagcaa gcagcacagg acccacctaa tattgatggc cagtggtcgc
 301 aattgcctgg ctaattagct ggtgcaccat ctcatttcc gcgtggattt tagctttccc
 361 acgcgggtta ggtcccttgt gctcttcttt aattaaccga tc
```

*Jla*244 HQ613864

```
   1 atttaaaatt tctcctataa aagtgttttc atttgggcgc tgaacaactt tgtcaaaaga
  61 aaatggaagg caaaaggctt ctttaaatgg ggacgacgcc aagccttgaa agtggacggg
 121 gattatatat taaaaaaaaa agagagagag agagagagag agagaaagga ggaattgatg
 181 gcctcaattt gcccttccca cttgaataga gtcatgatgg tggctaggac taaggaagaa
 241 ggtgctacga ctccctacat tagctctgac tttcggctaa gtgctgggga catccatctc
 301 gacccatcat gcgactcgtg gctgagagaa agtcagagga tc
```

*Jla*272 HQ613865

```
   1 gagaaagaga ggtttccttc tgcaagacaa acttgtcagg aatgattagg aggaggtgcg
  61 cccataggtc gtacggccag agcgtgcgga cattgctttc cgggggctcg tccttctcaa
 121 tagcaaattg gccaggaaaa agaaaataaa aaaaaaagaa agaaaatgac gtt[ctctctc]
 181 [tctctctctc tctctctctc tct]gcctcaa tttcttttga actttgaatt tactcactga
 241 tc
```

*Jla*273 HQ613866

```
   1 attgcggtta aaattggggt tcaaccagac taaaatccag cggaccgccc ttaactaact
  61 cgctttggtt aatccccaaa accccaccta ccggattccc cgcaaacgaa tcaccgaatt
 121 ctcaagcaaa atgtacgcgt gtgcgt[gaga gagagagaga gagagagaga gagagagaga]
 181 [gagaga]gtga ccgagaagga aggaatgaat gagtgaacac tgaccgtcgg atgtaggagc
 241 tgatgaaaga atttagaagc atcgaacagc accactctcg ggacaatgaa caaatggaca
 301 atgatggaac atcaatgatg agaacggaca aacagatacc aagaacaaaa taatggatcg
 361 atcgttcctt ggcagaccat ctgcgccaaa attcgcgaat ttatagtggc cctttctcta
 421 tcagtgctcc gggacggttt cggctgttat attagtgcgg cggactgggg gacgcccgtt
 481 cgatggtcct ctgttttccc catctatctc ttttccccac ctctctctct ctctctctct
 541 ctctctctct ga
```

*Jla*282 HQ613867

```
   1 catccgcggc aaacgcgagc ccccataaat aatttgagac ggagccccat tccactgggt
  61 ttccccacct ttcggtatct ccgatagtca ctttaaccaa acaacggttg acaagtggca
 121 agatgtcggc tggcctcgcc aaatcatttg agcatgcata tggatgctcc tatgacagag
 181 aacccgttca ccttttaat aggctatatc atagaa[gaga gagagagaga gagagagaga]
 241 [gagagagaga gagaga]ggga gaggccgtgt tattgccgtg aggtgatggt gggagagaga
 301 cagcatggta atacgtggag gaaggtgaga gaaactatca caatcaaaat aattttatag
 361 gtagattata tagatgacaa gtatctctgg tcataattac tcaaatattt cattaaaact
 421 tattcatgag agacttgaag aaagaggcgt cttctcctct tgatc
```

*Jla*299 HQ613868

```
   1 tggaagctcg catccagaga agagaatgta gatggtgcgg acgactctct tgacgtcaaa
  61 tcgaccgaaa tcggcctccg ctcgccctgg ctcctgaagt a[gagagagag agagagagag]
 121 [agagagagag agagagagag a]acagaaaca agagaaatta aaattctaat gctttgaagc
 181 gtaggttgac aggaacgggc tcgatc
```

*Jla*305 HQ613869

```
   1 ggcatcgggg ggaagtctat ttgacatttc tgcggattgt tacgtgcata agcgaggcgc
  61 cagaggctgt cgagaataga gatatagctt ctttgaagcg cctccctttc ctgctacttg
 121 tgcttgctta ctggactggc ttggcgagct ccggaaagaa gatcgatcta tttgaatcca
 181 acggctgata tctcattatc aagacccata tctcgacgcc atctatatat ctcccaccac
 241 ccccgctcct ctcctttgcg aaaagcttcc gccatactaa acggggaggg gaaggggg[ga]
 301 [gagagagaga gagagagaga gaga]ggggaa tcggacggct gtaatctctc aaagggtgga
 361 acggatggcg acgatgggtt accggagctc gagcagca
```

*Jla*328 HQ613870

```
   1 ctccctttga tgttttcgag agctaagaaa cgtaagaaaa gatactaaaa tttgaaaaca
  61 tgccatgata ttcagtttcc tcaaaagaaa aggcctttct tttgggaaag ttaggccaaa
 121 ggtgtgaagg cccgggccac accacttgac cggtaagcca [ctctctctct ctctctctct]
 181 [ctctctctct ctctctctct ctctctctct] ttctcctttt gttttcatgg ttgcccatca
 241 ctacattcgc caaaaatcga gatc
```

*Jla*330 HQ613871

```
   1 ccagcccccc atttctctct ctctctcggc gctcgtctca atggcttatc cgccatggcc
  61 cactctcctc tccctcctct gcgttttctt catcacctct ctatctctcc ctctccctct
 121 cgggtccgtt tcccacccat ccattaattc cttatccttc at[ctctctct ctctctctct]
 181 [ctctctctct ctctctctct ctctctctct] cattctctgt atgtgtgtct gactgcattc
 241 cttgacgatt tcagtgcaga tataaaggca ctgaccgttg gagatgagat c
```

*Jla*332 HQ613872

```
   1 ctaaaaagaa agttattgct ctcaaggcct ccggacaaaa aagtgaagat gagagtagtg
  61 aggaagagag tgaggaggat gaagatgtgg ccctcatagc tagaaggttt agaaaattca
 121 tgaggaaaag aaagccacac ttcaagagga gaatatctag aggtgaacaa gaaaaagaaa
 181 gagataatag ataaggagaa agaaaaagaa caagttttgt gctatgagtg taaaaagata
 241 gggcattata gagatcagct ccgatggttc cggacggaaa cggtagccac gaggcaatcc
 301 catcgcacga acctttctct ctctctctct ctctctctct ctctgcctcc ccctcaaagg
 361 catgaaagac tctcctatat gagcattata tatatatata tatgcataga ctgcctcctc
 421 cttcccccct ccacccccttc ccgggggaga agttggggag ggagctcggg gaggtactac
 481 aagtcaagaa gggcggcatc catgtgaaag ataatagaag gacagggg
```

APPENDIX B: Paper published in American Journal of Botany Primer Notes & Protocol in the Plant Sciences.

American Journal of Botany

American Journal of Botany: e117–e119. 2011.

AJB Primer Notes & Protocols in the Plant Sciences

Isolation and characterization of microsatellite loci in an endangered palm, *Johannesteijsmannia lanceolata* (Arecaceae)[1]

Cheng Choon Ang[2], Soon Leong Lee[3], Chai Ting Lee[3], Lee Hong Tnah[3], Rozainah Mohamad Zakaria[2], and Ching Ching Ng[2,4]

[2]Institute of Biological Sciences, Faculty of Science, University of Malaya 50603 Kuala Lumpur, Malaysia; and [3]Genetic Laboratory, Forest Research Institute Malaysia 52109 Kepong, Selangor, Malaysia

- *Premise of the study:* Microsatellite markers were developed for *Johannesteijsmannia lanceolata* to assess the genetic diversity and mating system of this alarmingly endangered species.
- *Methods and Results:* A total of 31 polymorphic microsatellite markers were developed for *J. lanceolata* using the enrichment protocol. These markers were screened on 24 samples from a natural population. The number of alleles ranged from four to 20, while the observed heterozygosity ranged from 0.391 to 1.000. The 31 loci were further tested for transferability on *J. altifrons*, *J. magnifica*, and *J. perakensis*. Generally, all loci showed positive amplifications in these three *Johannesteijsmannia* species, except *Jla*124 (*J. magnifica*) and *Jla*168b (*J. magnifica* and *J. perakensis*).
- *Conclusions:* These microsatellite markers could be employed to study the population genetics and mating system of *J. lanceolata* and other *Johannesteijsmannia* species.

Key words: cross-species amplification; DNA marker; enriched genomic library; short tandem repeat.

The palm family Arecaceae (or Palmae) includes about 2 364 species in 190 genera that are generally distributed throughout the tropics and subtropics of America and Asia (Govaerts and Dransfield, 2005). *Johannesteijsmannia lanceolata* J. Dransf. is an endemic palm species of Peninsular Malaysia with narrow distributions in Negeri Sembilan, Pahang, and Selangor. It is an understory palm that grows in well-drained soils near the steep river banks or hill slopes of primary tropical rainforests (Dransfield, 1972). *Johannesteijsmannia lanceolata* is regarded as a highly valuable ornamental palm due to its exotic, long, undivided, lanceolate leaves. *Johannesteijsmannia lanceolata* was listed as an endangered species in the 1997 IUCN red list of threatened plants due to critical population reduction (Walter and Gillett, 1998). To date, several phenological studies on flowering and vegetative growth have been conducted on *J. lanceolata* (Rozainah and Sinniah, 2006), but there is no genetic information published regarding this species. Therefore, understanding the population structure and genetic diversity using microsatellite markers is important to postulate scientific conservation and restoration strategies for the species.

Here we exemplify the isolation of 31 polymorphic microsatellites for *J. lanceolata* and performed cross-amplification tests in three other *Johannesteijsmannia* species (*J. altifrons* (Reichb. f. et Zoll.) H. E. Moore, *J. magnifica,* J. Dransf. and *J. perakensis* J. Dransf.).

[1]Manuscript received 2 December 2010; revision accepted 12 January 2011.

This research was funded by the Ministry of Agriculture and Agro-Based Industry, Malaysia, under the Science Fund (05-01-03-SF0199 and 05-03-10-SF1031) and postgraduate research fund (PS228/2009C). The research was conducted mainly in the Genetic Laboratory, Forest Research Institute Malaysia. The authors thank Dr. Kevin Ng Kit Siong, Dr. Ng Chin Hong, Chan Yoke Mui, Dr. Naoki Tani, and Yeoh Suat Hui for assistance in the laboratory and for valuable comments.

[4]Author for correspondence: ccng@um.edu.my

doi:10.3732/ajb.1000494

METHODS AND RESULTS

Total genomic DNA was extracted from leaf tissues of *J. lanceolata* using a modified cetyltrimethyl ammonium bromide (CTAB) method (Murray and Thompson, 1980). A CT-enriched library was constructed according to Lee et al. (2004). Approximately 5 µg of genomic DNA was digested with *Nde* II (Promega, Madison, Wisconsin, USA), and the DNA fragments of 300–1 000 bp were selected to be ligated with Sau3A1 cassettes (Takara, Otsu, Shiga, Japan). Nicks between DNA fragments and *Sau*3A1 cassettes were repaired after ligation using DNA polymerase I (Takara, Otsu, Shiga, Japan). Cassette-ligated DNA was enriched for microsatellite repeats by hybridization to 5′ biotinylated $(CT)_{15}$ probe and captured by streptavidin-coated magnetic beads (Promega, Madison, Wisconsin, USA). The selectively recovered fragments were amplified by polymerase chain reaction (PCR) with C1 cassette primers and digested with *Nde* II to remove the cassettes. These fragments were subsequently ligated into pUC118 *Bam*H1/BAP (Takara), transformed into Qiagen EZ competent cells (Qiagen Gmbh, Hilden, Germany), and plated on LB agar containing ampicillin (100 µg/mL), IPTG (50 µM), X-gal (80 µg/mL). Insert-containing clones were selected by blue/white screening. Circular DNA was amplified using Templiphi kit (Amersham Biosciences, Piscataway, New Jersey, USA) and sequenced using an ABI 3130xl Genetic Analyzer with BigDye Terminator v3.1 Cycle Sequencing Kits (Applied Biosystems, Foster City, California, USA) and M13 primer.

A total of 285 nonredundant sequences were obtained from 336 clones by the Cd-hit program, revealing 15.2% redundancy. A total of 353 microsatellites were identified from 226 clones using the MISA-MIcroSAtellite identification tool. Based on these sequences, a total of 111 primer pairs were designed using OLIGO 6 software (Molecular Biology Insight, Cascade, Colorado, USA). Four individuals of *J. lanceolata* from a natural population [Angsi Forest Reserve (2°43′4.50″N, 102°4′ 6.30″E), Negeri Sembilan, Malaysia] were selected for initial screening of these primer pairs. PCR amplifications were carried out in a 10-µL reaction volume with approximately 10 ng of template DNA, 50 mM

TABLE 1. Characteristics of the 31 polymorphic microsatellite markers developed in *Johannesteijsmannia lanceolata*, including forward (F) and reverse (R) primer sequences, repeat motifs, allele size ranges, annealing temperatures (T_a), and GenBank accession numbers. Significant departure from Hardy–Weinberg equilibrium after Bonferroni adjustment ($P < 0.001613$) is denoted with an asterisk.

Locus	Primer sequence (5′–3′)	Repeat	Allele size (bp)	T_a (°C)	GenBank Accession No.
*Jla*002	F: GGTGTGGTGCAAGGGAGTAT R: TCTCATCTACTTGGACGTCAGTGT	$(GA)_{16}$	201–233	45	HQ613842
*Jla*046a	F: TTGCCTATGTAAAATGTTAACTAA R: AGAGGGTGAGGGTTCGATT	$(CT)_{24}$	203–257	45	HQ613843
*Jla*048	F: TTCCCTTCTTTGCACGATTTCGAG R: GCATGGGAGGGTTGGTAGGGTTT	$(GA)_{15}$	140–154	45	HQ613844
*Jla*106	F: GCCCATGATTAGCTTTAACC R: TATGTTTGCCTCCCTCAGACTTGA	$(CT)_{20}$	228–254	55	HQ613845
*Jla*112	F: AATGTACTTTCCCTGAGCTCTACT R: ATCACTGCCAATCTTATCACC	$(CT)_{21}$	208–242	55	HQ613846
*Jla*117	F: TTGGCAATGGAGAGAACGAGAACA R: CCAAATAATTTCAAGCGTTCCTC	$(GA)_{13}$	243–261	45	HQ613847
*Jla*118	F: CACTCCAAGAAGAACCCCGATG R: ATCTTTCTCTTTTTGCTCCGCTCC	$(CT)_{22}$	139–171	50	HQ613848
*Jla*124	F: AGTGGCCTTTGAATTATGTC R: CCTCCATGGCTACAACAATTAGTT	$(CT)_{19}$	193–273	45	HQ613849
*Jla*130b*	F: CAACTTGGCCTATGCTCAC R: AGGGACCCATTAAGTAAAGCAACT	$(CT)_{24}$	113–133	45	HQ613850
*Jla*131	F: GGGACAATAAATCATGAACAATTA R: AGACATGCACATGCACACG	$(CT)_{28}$	77–99	45	HQ613851
*Jla*160	F: CCCTCCTCTCTGTCTCCCCATTTC R: AACTCCCCAAACAGCACGCAATC	$(CT)_{15}$	76–90	45	HQ613852
*Jla*162	F: CCCACAGCCCCACTGATTGATT R: TGTGTGTACAGTCCGTGCCGTCAC	$(CT)_{25}$	179–201	45	HQ613853
*Jla*168b	F: GAGAGTGCCCACTTGAGTCA R: GTCCAACAAGCATCAGACCCTTAT	$(GA)_{25}$	180–228	45	HQ613854
*Jla*174	F: CAGAGGTAATGCAAAATCAACCCC R: TTATTGCAGATATGGCCCTATT	$(CT)_{19}$	89–109	45	HQ613855
*Jla*186	F: ACCAAACACGGACATTCTCAAGAT R: CGGCTCATTGGAATATGTCCT	$(CT)_{17}$	160–190	45	HQ613856
*Jla*187b	F: GTGCATGCGTGTGCGTGCG R: TTTGGTACTAATGTCTCCGGTTGA	$(GA)_{18}$	141–183	45	HQ613857
*Jla*192	F: TTACATGTTTTGGTATCAGGTTTT R: CAACACGAGGAATAGGCTA	$(CT)_{21}$	138–179	45	HQ613858
*Jla*196a	F: AAATGTACCCAAATCACAACTCAA R: TACAAGCGAGTGCACGGATA	$(CT)_{27}$	169–209	45	HQ613859
*Jla*198	F: CATCAAAATCAAGAATCTTATGGC R: CGGATGAATAATACGTGCAAGG	$(GA)_{21}$	251–291	45	HQ613860
*Jla*210	F: TGTCCTCTACTTCCCACAACAATC R: CGCAATGTCGTGGTCTACAGT	$(CT)_{22}$	149–195	45	HQ613861
*Jla*236	F: AGGCCTGATTCCACTGCACTAGAT R: GCATGAGAAAACGAAAGGGGTATA	$(CT)_{20}$	75–93	45	HQ613862
*Jla*238	F: CACAAGATTTTTAGCAGGTAGGAT R: ACTGGCCATCAATATTAGG	$(CT)_{20}$	176–200	45	HQ613863
*Jla*244	F: GCTTCTTTAAATGGGGACGAC R: CCGAAAGTCAGAGCTAATGTAGGG	$(GA)_{12}$	190–212	45	HQ613864
*Jla*272	F: GGGCTCGTCCTTCTCAATAG R: CAAAGTTCAAAAGAAATTGAGGCA	$(CT)_{15}$	98–121	45	HQ613865
*Jla*273a	F: GCCCTTAACTAACTCGCTTTG R: CATCAGCTCCTACATCCGACGGTC	$(GA)_{20}$	181–203	45	HQ613866
*Jla*282	F: CCTATGACAGAGAACCCGTTCACC R: ACCATCACCTCACGGCAATAACAC	$(GA)_{20}$	114–124	45	HQ613867
*Jla*299	F: TGTAGATGGTGCGGACGACT R: CGTTCCTGTCAACCTACGCTTCAA	$(GA)_{20}$	164–190	45	HQ613868
*Jla*305	F: GAAAAGCTTCCGCCATACTAAAC R: GAGAGATTACAGCCGTCCGATTCC	$(GA)_{13}$	78–100	45	HQ613869
*Jla*328	F: CCACACCACTTGACCGGTAAGC R: CGATTTTTGGCGAATGTAGTGATG	$(CT)_{25}$	96–144	55	HQ613870
*Jla*330b	F: GTTTCCCACCCATCCATTAATTC R: CCAACGGTCAGTGCCTTTATATCT	$(CT)_{24}$	115–153	45	HQ613871
*Jla*332b	F: CCCTCAAAGGCATGAAAGACTCTC R: GCCCTTCTTGACTTGTAGTACCTC	$(TA)_8$	139–143	45	HQ613872

KCl, 20 mM Tris-HCl (pH 8.0), 1.5 mM $MgCl_2$, 0.2 μM each primer, 0.2 mM dNTP, and 0.5 Unit GoTaq Flexi DNA polymerase (Promega, Madison, Wisconsin, USA). The PCR thermal profile was: initial denaturation at 94°C (5 min), 35 cycles of 94°C (30 s), 45°C (30 s), 72°C (30 s) and followed by final extension at 72°C (7 min). The PCR products were electrophoresed on 2% agarose gel. Out of 111 primer pairs, 31 pairs that yielded specific amplicons of expected fragment size without multiple bands were labeled with fluorescent dyes (6-FAM or HEX) at the forward primers. Twenty-four samples from the same population were then screened using similar PCR conditions with modification of the annealing temperatures for four of the loci (Table 1) to reduce nonspecific amplifications. The PCR products were subjected to fragment analysis using ABI 3130xl Genetic Analyzer with ROX 400 (Applied Biosystems) as the internal size standard in assigning allele sizes and further scored by using GeneMapper v 4.0 (Applied Biosystems).

All the 31 primer pairs tested were polymorphic (Table 1). Characterization of these polymorphic loci was further performed using CERVUS version 3.0.3 (Kalinowski et al., 2007). The number of alleles per locus (A) observed from the 24 samples ranged from four to 20 with an average of 10. The observed heterozygosity (H_O) and expected heterozygosity (H_E) ranged from 0.391 to 1.000 and from 0.542 to 0.957, respectively. Polymorphic information content (PIC) ranged from 0.491 to 0.932 (Table 2).

Exact tests for Hardy–Weinberg equilibrium and linkage disequilibrium were carried out using GenePop version 4.0.10 (Rousset, 2008), and Bonferroni correction for multiple tests was implemented. A significant departure from the Hardy–Weinberg equilibrium was detected in *Jla*130b ($P < 0.001613$). Microchecker (van Oosterhout et al., 2004) analysis suggested that *Jla*118 was likely to contain null alleles at 99% confidence interval. No significant linkage disequilibrium was detected between any pair of loci ($P < 0.000108$).

Subsequently, the 31 loci were subjected to the Mendelian mode of inheritance test using a half-sib family consisting of 15 progenies. The results showed that the amplified fragments of all loci segregated in a codominant manner, and each of the progeny possessed at least one maternal allele, which supported the postulation of a single-locus mode of inheritance. The 31 loci were further tested for transferability on *J. altifrons* from Endau-Rompin National Park, Johor (2°26′20.09″N, 103°16′22.29″E), *J. magnifica* from Jeram Toi Recreational Forest, Negeri Sembilan (2°51′52.98″N, 102°0′50.34″E) and *J. perakensis* from Kledang Sayong Forest Reserve, Perak (4°42′57.24″N, 100°58′2.16″E) using four individuals per species. Amplifications were considered positive when the primers yielded specific PCR products of expected size without multiple bands. In general, all loci showed positive amplifications in these three *Johannesteijsmannia* species, except *Jla*124 (*J. magnifica*) and *Jla*168b (*J. magnifica* and *J. perakensis*).

TABLE 2. Genetic diversity parameters of the 31 polymorphic microsatellite loci based on 24 samples of *Johannesteijsmannia lanceolata*. Number of alleles per locus (A), observed (H_O) and expected (H_E) heterozygosities, and polymorphic information content (PIC) are given for each locus.

Locus	A	H_O	H_E	PIC
*Jla*002	13	0.913	0.902	0.872
*Jla*046a	19	0.864	0.911	0.881
*Jla*048	7	0.667	0.803	0.757
*Jla*106	10	0.917	0.882	0.849
*Jla*112	13	0.913	0.899	0.868
*Jla*117	7	0.739	0.829	0.784
*Jla*118	11	0.636	0.877	0.842
*Jla*124	11	0.682	0.782	0.740
*Jla*130b	9	0.708	0.849	0.811
*Jla*131	10	0.667	0.793	0.752
*Jla*160	5	0.542	0.621	0.531
*Jla*162	9	0.792	0.815	0.771
*Jla*168b	18	0.833	0.941	0.916
*Jla*174	7	0.708	0.775	0.727
*Jla*186	11	0.833	0.845	0.811
*Jla*187b	7	0.739	0.853	0.813
*Jla*192	11	1.000	0.905	0.875
*Jla*196a	20	0.913	0.957	0.932
*Jla*198	13	0.958	0.902	0.872
*Jla*210	12	0.792	0.881	0.848
*Jla*236	7	0.792	0.762	0.705
*Jla*238	12	0.875	0.905	0.876
*Jla*244	11	0.739	0.671	0.641
*Jla*272	9	0.667	0.841	0.801
*Jla*273a	11	0.957	0.909	0.879
*Jla*282	6	0.833	0.809	0.763
*Jla*299	8	0.833	0.833	0.793
*Jla*305	8	0.652	0.794	0.746
*Jla*328	14	0.917	0.834	0.803
*Jla*330b	10	0.870	0.786	0.746
*Jla*332b	4	0.391	0.542	0.491

CONCLUSIONS

These 31 newly developed microsatellite markers exhibited a high level of polymorphism that could be used in the assessment of the genetic diversity and mating system of *J. lanceolata* across its distribution in Peninsular Malaysia. Better understanding of the species' mating system and population genetic structure is crucial to set conservation strategies and priorities of this endangered species.

LITERATURE CITED

DRANSFIELD, J. 1972. The genus *Johannesteijsmannia* H. E. Moore Jr. Garden's Bulletin vol. 26, part 1, 63–83, Government Printing Office, Singapore.

GOVAERTS, R., AND J. DRANSFIELD. 2005. World checklist of palms. Royal Botanic Gardens, Kew, Richmond, Surrey, UK.

KALINOWSKI, S. T., M. L. TAPER, AND T. C. MARSHALL. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* 16: 1099–1106.

LEE, S. L., N. TANI, K. K. S. NG, AND Y. TSUMURA. 2004. Isolation and characterization of 20 microsatellite loci for an important tropical tree *Shorea leprosula* (Dipterocarpaceae) and their applicability to *S. parvifolia*. *Molecular Ecology Notes* 4: 222–225.

MURRAY, M. G., AND W. F. THOMPSON. 1980. Rapid isolation of high molecular-weight plant DNA. *Nucleic Acids Research* 8: 4321–4325.

ROUSSET, F. 2008. Genepop' 007: A complete re-implementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources* 8: 103–106.

ROZAINAH, M. Z., AND U. R. SINNIAH. 2006. Flowering event of *Johannesteijsmannia lanceolata*, an understorey palm in the Angsi Forest Reserve, Malaysia. *Malaysian Journal of Science* 25: 65–68.

VAN OOSTERHOUT, C., W. F. HUTCHINSON, D. P. M. WILLS, AND P. SHIPLEY. 2004. Micro-checker: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 4: 535–538.

WALTER, K. S., AND H. J. GILLETT. [eds.] 1998. 1997 IUCN Red List of Threatened Plants. Compiled by the World Conservation Monitoring Centre. IUCN-The World Conservation Union, Gland, Switzerland and Cambridge, UK.