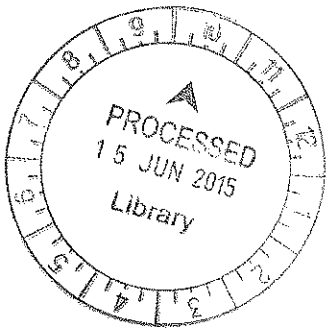
COMPARING *matK* AND *rbcL* SEQUENCES AMONG *Jasminum grandiflorum, Jasminum sambac AND Jasminum nitidum*

LAM CHI YAN

DISSERTATION SUBMITTED IN PARTIAL FULFILMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
BACHELOR OF BIOTECHNOLOGY (HONOURS)



FACULTY OF SCIENCE, TECHNOLOGY,
ENGINEERING AND MATHEMATICS
INTI INTERNATIONAL UNIVERSITY
PUTRA NILAI, MALAYSIA

2015

TABLE OF CONTENT

NON-PLAGIARISM DECLARATION

By this letter I declare that I have written this dissertation completely by myself, and that I have used no other sources or resources than the ones mentioned.

I have indicated all quotes and citations that were literally taken from publications, or that were in close accordance with the meaning of those publications, as such. All sources and other resources used are stated in the references.

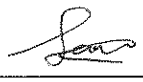
Moreover I have not handed in a dissertation similar in contents elsewhere.

In case of proof that the dissertation has not been constructed in accordance with this declaration, the Faculty of Science, Technology, Engineering and Mathematics has the right to consider the research dissertation as a deliberate act that has been aimed at making correct judgment of the candidate's expertise, insights and skills impossible.

I acknowledge that the assessor of this item may, for the purpose of assessing this item,

- reproduce this assessment item and provide a copy to another member of the University; and/or,
- communicate a copy of this assessment item to a plagiarism checking service (which may then retain a copy of the assessment item on its database for the purpose of future plagiarism checking).

In case of plagiarism the examiner has the right to fail me and take action as prescribed by the rules regarding Academic Misconduct practiced by INTI International University.

Lam Chi Yan	
Name	Signature
I12001211	8/5/2015
I.D.Number	Date

ACKNOWLEDMENT

First and foremost, I would like to express my sincere gratitude to my supervisor, Dr. Geeta, who has supported me throughout my thesis with her patience. Her guidance helped me in the time of research and writing of this thesis, without my supervisor this thesis would not been completed or written.

I thank my lab partner, Chang Kek Ning, who gave me support during the research. When we faced the difficulties, we encouraged each other. Furthermore, we helped each other in the lab. Thus, my research project was done quickly and smoothly. Besides that, I thank my coursemates who gave me support to complete this thesis. Although it was difficult to complete, we encouraged and supported each other to complete the writing.

Last but not the least; I would like to thank my family for supporting me spiritually throughout my life.

ABSTRACT

DNA barcodes are used to assign samples of unknown origin to specific taxonomic groups i.e., short DNA sequences (600-800bp) of a standard segment for species identification. In my study, I used the *matK* and *rbcL* barcode sequences to identify the *Jasminum grandiflorum, Jasminum sambac* and *Jasminum nitidum*. In these species, Edwards's method was able to provide of sufficient quality and quantity of DNA for PCR. In my results, the *matK* gene showed ease of amplification compared to *rbcL* gene. The second trial of *matK* with diluted DNA samples show higher success rate. For species discrimination, the neighbor-joining tree was used to construct phylogenetic trees based on T92 model as it was the best fit sequence model based on *matK* and *rbcL* gene. Based on the *matK* gene, *J. grandiflorum* is more closely related to *J. sambac* (Distance= 0.010) then *J. nitidum* (Distance= 0.020); while based on the *rbcL* gene, *J. sambac* is more closely related to *J. nitidum* (Distance= 0.007) then *J. grandiflorum* (Distance= 0.018). Although the tree topology based on *matK* and *rbcL* genes were different, the individuals of the three species grouped together. Hence, *J. sambac*, *J. grandiflorum* and *J. nitidum* can be identified using *matK* and *rbcL* gene. Further research to identify and compare with other *Jasminum* species is needed.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

%	Percentage
µL	micro litre
°C	Degree Celsius
V	voltage
dH_2O	Distilled water
bp	Base pair
AIC	Akaike information criterion
BIC	Bayesian information criterion
BLAST	Basic local alignment search tool
BLASTN	Nucleotide blast programme
CO1	Cytochrome c oxidase subunit 1
CBOL	Consortium for the barcode of life
CTAB	Cetyltrimethyl ammonium bromide
DNA	Deoxyribonucleic acid
EDTA	Ethylene diamine tetraacetic acid
g	gram
HCl	Hydrochloric acid
HKY	Hasegawa-Kishino-Yano
ITS	Internal transcribed spacer
ITS2	Internal transcribed spacer 2
kb	kilobases
K2P	Kimura 2-parameter
M	Mol per litre

mL	millilitre
matK	Megakaryocyte-associated tyrosine-protein kinase
MEGA	Molecular evolutionary genetics analysis
MUSCLE	Multiple sequence comparison by Log-expectation
mw	Molecular weight
NaCl	Sodium chloride
NCBI	National center for biotechnology information
NJ	Neighbor-joining
PCR	Polymerase chain reaction
pH	Potential of hydrogen
rbcL	Ribulose-biphosphate carboxylase gene
RNase A	Ribonuclease A
rpm	Revolutions per minute
SDS	Sodium dodecyl sulfate
T92	Tamura,1992
TBE buffer	Tris-borate-EDTA
TE buffer	Tris/ EDTA
TER buffer	Tris/ EDTA buffer with RNase A
UPGMA	Unweighted pair-group method using arithmetic averages

1.0 CHAPTER 1:

INTRODUCTION

DNA barcode is mainly used to assign samples of unknown origin to specific taxonomic groups and is important for discovery of land plants. The evolutionary distances and relationships among species can be distinguished by DNA barcode faster and more efficiently (Stoeckle, Waggoner, & Ausubel, 2005). Additionally, species that look alike can be distinguished, and even incomplete samples consisting of parts of a species can be identified (Stoeckle, Waggoner, & Ausubel, 2004).

DNA barcoding could be used to authenticate different *Jasminum* species or varieties used for medicinal purposes. *Jasminum* is generally known for its medicinal importance as it is used to treat many diseases such as skin disease and amenorrhea and as an antidepressant, anti-inflammatory and antioxidant (Jain, Sharma, Kumar & Sharma, 2011). And *Jasminum* as one of the largest genus in the *Oleaceae* family consisting of more than 200 species (Mahmood, Hafiz, Abbasi & Faheem, 2013/ Shekhar, Sriram & Prasad, 2013) may pose difficulties in identification. The *Jasminum* species that will be used in this research are *Jasminum grandiflorum, Jasminum sambac and Jasminum nitidum.*

To select a standard barcode for *Jasminum* species is challenging. This is because every barcode has different strengths among different plant species (Fazekas, Kuzmina & Newmaster, 2012). For this research, megakaryocyte-associated tyrosine-protein kinase (*matK*) and ribulose-biphosphate carboxylase (*rbcL*) gene barcode were selected to test as barcode identifiers in *Jasminum*. Consortium of the barcode of life (2009), approved these two loci as the barcode regions for land plants, and there are universal primers available for *matK* and *rbcL*. However, CBOL (2009) has indicated that additional works is needed on *matK* since *matK* primers are not easy to amplify in some species.

For this project, the aim is to obtain and compare the DNA sequences among *Jasminum grandiflorum*, *Jasminum sambac*, and *Jasminum nitidum* for the *matK* and

rbcL gene. These sequences will then be analyzed for their suitability to distinguish the *Jasminum* species.

2.0 CHAPTER 2:

LITERATURE REVIEW

2.1 *Jasminum* SPECIES

Species within the genus, *Jasminum* which are angiosperm were selected to be studied in terms of their DNA sequence. *Jasminum* is also known as jasmine. It is mostly planted in houses for ornamental purposes. In addition, it also plays an important role in medical and pharmaceutical industries (Shekhar, Sriram & Prasad, 2013). The essential oil that is extracted from *Jasminum* flowers can be used for a wide range of purposes. It is not only used in the medicinal industry, but also used in the cosmetic industry. Other than India and China, Malaysia is one of the regions that *Jasminum* species is mainly distributed (Singh, 2006). However, there is little work on the study of DNA sequences of *Jasminum*. Most of the literatures studying genetic relationship among *Jasminum* species use other methods. For example, Mahmood, Hafiz, Abbasi and Faheem (2013) used randomly amplified polymorphic DNA to study the genetic relationship among *Jasminum* species. They successfully distinguish and classified the genetic characteristics of a few *Jasminum* species such as *J. grandiflorum*, *J. sambac* and *J. humile* into different groups. However, there is further research needed for DNA Barcoding to identify other *Jasminum* species.

2.2 DNA BARCODE FOR PLANTS

DNA barcode is widely used to identify species. Most literature has indicated that DNA Barcoding in plants is more difficult than in animals (Janzen, 2009). This is because no gene has so far been useful as a plant barcode for all plant species. For example, the mitochondrial gene cytochromes c oxidase subunit 1 (*CO1*) barcode can work well for animals, but does not work well in plants due to its lower rate of change in gene sequence (Janzen, 2009).

There are various other genes that are suggested for a plant barcode. For example, there are coding genes such as *matK* and *rbcL* and the non-coding spacers

such as *trnH-psbA*. Different genes have different strengths and weaknesses. Thus, the gene that works well in angiosperm species may not work well in gymnosperm species. Among the plant barcodes, Janzen (2009) has proposed that *rbcL* and *matK* are the potential barcodes for plant DNA Barcoding. In addition, *rbcL* and *matK* were also the majority choice among the 2-locus barcode combination (Janzen, 2009). This is because *matK* has high levels of discrimination and resolution among angiosperm species and *rbcL* has high universality in discrimination of species. Hence, *matK* and *rbcL* were selected to study the DNA sequence of *Jasminum* in this research.

2.3 CHALLENGES FOR DNA BARCODING

Although *matK* and *rbcL* barcode were recommended as a potential barcode in plant DNA Barcoding, some literatures have shown that there are weaknesses in using *matK* and *rbcL* barcode. For instance, Olivar, Brillantes, Rubite and Alejandro (2014) and Li et al (2010) concluded that *matK* exhibited low universality due to low polymerase chain reaction (PCR) amplification and sequencing whereas rbcL was weak in discrimination. According to Fazekas, Kuzmina, and Newmaster (2012), *matK* gene exhibited low universality because it is approximately 300 bp longer than *rbcL*; hence, *matK* gene is more sensitive to DNA degradation.

Furthermore, Pettengill and Nell (2010) concluded that *matK* and *rbcL* was the worst-performing barcode to distinguish the species in genus *Agalinis* because the barcodes regularly required multiple attempts at PCR and sequencing to obtain a high quality sequence from samples. In addition, most of the *matK* and *rbcL* failed in PCR amplification (Bafeel et al, 2011). The research of Li et al (2010) showed the same problem as Pettengill and Nell (2011) which is that the *matK* gene showed the lowest efficiencies in identification although they worked on other species, i.e., of *Taxillus chinenesis* species and its related parasitic plant *Loranthus*. Among other barcodes they tested *matK* performed poorly due to its failure in PCR amplification and sequencing.

However, the *rbcL* gene, in the research of Olivar, Brillantes, Rubite and Alejandro (2014) does not show any difficulties in PCR amplification and sequencing for distinguishing *Ficus L.* species. However, it could not resolve the *Ficus L.* species.

This was because it exhibited low inter-specific divergence and low sequence divergence (Li et al, 2010). Although *rbcL* has proven that it was easily sequenced and aligned with plant groups such as mosses, ferns and angiosperms in research, it was difficult to group the species. Thus, it will be interesting to see if DNA barcode identification of *Jasminum* species with *matK* and *rbcL* gene is possible.

2.4 WAYS TO OVERCOME PROBLEMS

As most literature stated that *matK* and *rbcL* has difficulties in PCR amplification and sequencing, further modification needs to be carried out to reduce the failure rate. For example Olivar, Brillantes, Rubite and Alejandro (2014) used *rbcL*, *ITS* and *trnH-psbA* to distinguish *Ficus L.* species. In their first trial in amplification, there were difficulties for carrying out the amplification of pure DNA. Thus, they made a modification for their second trial which is to dilute the DNA samples (1:10). Although the success rate of *rbcL* couldn't achieve 100 % in the second trial, it was way better than the first trial.

On top of that, Yu, Xue and Zhou (2011) have proposed a new *matK* primer to overcome the low universality problem. They evaluated the whole *matK* barcode region to find a region of 600-800 bp which was highly variable. This was because the fragment of 600-800 bp was usually sufficient in DNA Barcoding. In their result, the new *matK* barcode showed a strong amplification and sequencing successes. Hence, the *matK* loci amplified with the new primers are highly successful and can be used to test angiosperm samples.

Other than that, Fu, Jiang and Fu (2011) recommended the use of multilocus barcoding system instead of a single locus barcoding system. They found that a multilocus DNA barcode can perform better than a single locus. This is because the combination of barcode provides enough nucleotide variation to discriminate the species. For instance, the combination of *matK*, *rbcL* and *ITS* barcode was recommended in identification of species within *Tetrastigma* (Fu, Jiang & Fu, 2011) and the combination of *matK*, *rbcL* and *ITS2* region were recommended in identification of the species of *Gossypium* (Ashfaq, Asif, Anjum & Zafar, 2013). They established that the combination of barcodes were the best barcode based system for