DNA BARCODING OF PLANTS WITHIN THE MALVACEAE FAMILY

YAP LE XIAN

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

FACULTY OF HEALTH AND LIFE SCIENCES
INTI INTERNATIONAL UNIVERSITY
PUTRA NILAI, MALAYSIA

2017

NON-PLAGIARISM DECLARATION

By this letter I declare that I have written this thesis 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 thesis similar in contents elsewhere.

In case of proof that the thesis has not been constructed in accordance with this declaration, the Faculty of Health and Life Sciences has the right to consider the research thesis 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.

YAP LE XIAN
Name

Signature

I15008525
I.D.Number

15-12-2017
Date

Abstract

Nowadays, demand for medical plants has increased due to some side effects in synthetic drugs. This has led to some companies using plants with similar appearance as a replacement to make herbal products, and this has reduced the effect and quality of the drug. To solve this problem, a strong and reliable authentication method is required in order to differentiate the real from the fake medical plants. DNA barcoding is one of the suitable methods that can be used to identify the species of plant. It has a very high reproducibility with medium cost and ease of use in differentiation of plants. In this study, *Hibiscus rosa-sinensis* (with a single layer of white petals), *H. cannabinus L* and *Malvaviscus arboreus var drummondii* all of which belong to *Malvaceae* family and are said to have medical functions, were used. The DNA barcode markers that were used in this study to define the ability in differentiation of plant species are *matK* and *trnH-psbA*. The method of extracting DNA is an adaptation of a standard protocol which was established by Edwards, Johnstone & Thompson in 1991, and was successful in the Malvaceae inspite of their high polysaccharide content. Amplification of the *matK* and *trnH-psbA* sequence was successful and good quality sequences were obtained. While BLAST resulted in some ambiguities in identification, but the phylogenetic tree based on both *matK* and *trnH-psbA* clearly clustered samples of each species into the same clade with high bootstrap value of over 70%. However, when including the downloaded sequence of related species, only *Hibiscus rosa-sinensis* can clearly clustered into the same clade. Therefore, both the *matK* and *trnH-psbA* have the potential to differentiate each species of plant from others. Further study should be done by using larger samples size from different geographical areas and using others marker (such as nuclear marker).

LIST OF TABLE

Table		Page
2.5.1	Differences between the DNA-barcoding and other DNA-based Methods	6
3.4.1	Primer Sequence of *matK* and *psbA-trnH* used for amplification	12
3.4.2	Reagents present in the PCR mix used for PCR reaction	13
3.4.3	PCR amplification conditions for *matK* and *trnH-psbA* marker	13
4.4.1	The average quality (QV) and mismatch ratio of each sample.	21
4.7.1.1	The BLAST result of *trnH-psbA* marker sequences.	23
4.7.2.1	The BLAST result of *matK* marker sequences.	24
4.8.1.1	The first three best model to use for *matK* and *trnH-psbA* to build a phylogenetic tree.	25

LIST OF FIGURES

Figure		Page
3.3.1	Resulting bands for Vivantis (VC 1kb) DNA ladder on 1.0% agarose gel and their corresponding molecular weight.	11
4.1.1	Flower of each species of plant	17
4.2.1	Agarose gel electrophoresis (1.2% agarose) of DNA extracted from 3 variant species of plant	18
4.3.1	Agarose (1.2%) gel electrophoresis of *matK* amplified DNA product from 3 variant species of plant	19
4.3.2	Agarose (1.2%) gel electrophoresis of second round trnH-psbA amplified DNA product from 3 variant species of plant	20
4.9.1.1	Maximum likelihood tree for *matK* of 3 variant species of plant	26
4.9.1.2	Neighbor joining tree for *matK* of 3 variant species of plant	27
4.9.2.1	Maximum likelihood tree for *trnH-psbA* of 3 variant species of plant	28
4.9.2.2	Neighbor-joining tree for *trnH-psbA* of 3 variant species of plant	28
4.10.1	Maximum Likelihood tree for *matK* of variant plant	30

LIST OF ABBREVIATIONS

%	percentage
°C	degree Celsius
AFLP	amplified fragment length polymorphism
BLAST	basic local alignment search tool
Bp	base pair
CBOL	consortium of barcode of life
COI	Cytochrome Oxidase I
DNA	deoxyribonucleic acid
EDTA	Ethylene diamine tetraacetic acid
HCl	Hydrochloric acid
ITS	internal transcribed spacer
ITS2	internal transcribed spacer 2
M	molar
matK	Megakaryocyte tyrosine-protein kinase
MEGA 7.0	molecular evolutionary genetics analysis 7.0
mg/L	milligram per liter
NaCl	Sodium Chloride
NaOH	Sodium hydroxide
NCBI	national center for biotechnology information
psbA	photosystem II protein D1
RAPD	random amplified polymorphic DNA
rbcL	ribulose bisphosphate carboxylase
RNA	ribonucleic acid
RNase A	ribonuclease A
rpm	revolution per minute
SDS	Sodium dodecyl sulfate
TBE	Tris/Borate/EDTA
TE	Tris-EDTA
trnH	tRNA histidine
UV	ultraviolet light
v/v	volume per volume
w/v	weight per volume

Table of Content

	PAGE
DECLARATION	**ii**
ABSTRACT	**iii**
LIST OF TABLE	**iv**
LIST OF FIGURES	**v**
LIST OF ABBREVIATIONS	**vi**
TABLE CONTENT	**vii-ix**
CHAPTER	
1. INTRODUCTION	**1-2**
2. **LITERATURE REVIEW**	**3-8**
2.1 Medicinal Plants ----Malvaceae FAMILY	3
2.2 Introduction of Two Genera within Malvaceae	
2.2.1 *Hibiscus rosa-sinensis*	3
2.2.2 *Malvaviscus arboreus var drummondii*	4
2.2.3 *Hibiscus cannabinus L*	4
2.3 Authentication of Medical Herbs	4-5
2.4 Traditional Approach--- Morphology-based authentication	5
2.5 Alternative Approach---DNA-based Method	5-6
2.6 DNA Barcoding	6-7
2.7 Advantage of DNA-based Method Compared to Traditional Method	7
2.8 DNA Barcodes	7-8
3. METHODOLOGY	**9-16**
3.1 Sample collection	9
3.2 DNA Extraction	9-10
3.3 DNA Quality Assessment by Agarose gel electrophoresis	11
3.4 PCR Preparation	12-13
3.5 Preparation of Agarose Gel Electrophoresis to access the Quality of PCR Product	14
3.6 Sequencing and Sequence Quality Analysis	14
3.7 Sequence editing and BLAST Analysis	14-15
3.8 DNA Authentication	15

3.9 Multiple Sequence Alignment 15
3.10 Phylogenetic tree and DNA distance matrix construction 15-16
4. RESULT ANALYSIS 17-30
4.1 Sample Collection 17
4.2 DNA Extraction Analysis 17-18
4.3 PCR Product Analysis 19-20
4.3.1 Amplification using *matK* primer 20
4.3.2 Amplification using *trnH-psbA* primer 21
4.4 DNA Sequencing 21
4.5 Sequence Annotation (Dogma Analysis) 22
4.6 Fragment Length and Proportion of Bases (DAMBE Analysis) 22
4.7 BLAST Analysis
4.7.1 *trnH-psbA* primer 22-23
4.7.2 *matK* primer 23-24
4.8 Model Identification
4.8.1 Model for *matK and trnH-psbA* 24-25
4.9 Multiple sequence alignment and construct of phylogenetic tree
4.9.1 Construction of phylogenetic tree for *matK* 25-27
4.9.2 Construction of phylogenetic tree for *trnH-psbA* 28
4.10 Phylogenetic trees of different sample of plants of *matK* 29-30
5. DISCUSSION 31-38
5.1 DNA Extraction Analysis 31-32
5.2 PCR Product Analysis
5.2.1 Amplification using *matK* primer 33-34
5.2.2 Amplification using *trnH-psbA* primer 34
5.3 DNA Sequencing Analysis 34
5.4 Dogma Analysis 35
5.5 DAMBE Analysis 35
5.6 BLAST Analysis
5.6.1 *trnH-psbA* 35-36
5.6.2 *matK* 36
5.7 Model Identification 36-37

5.8 Multiple sequence alignment and construct of phylogenetic tree 37

5.9 Phylogenetic trees of *matK* for different species of plants of *matK* 37-38

6. LIMITATION AND SUGGESTION **39**

7. CONCLUSION **40**

REFERNCES **41-48**

APPENDIX 1 **49-51**

APPENDIX 2 **52-54**

APPENDIX 3 **55**

APPENDIX 4 **56-58**

APPENDIX 5 **59-61**

APPENDIX 6 **62-63**

CHAPTER 1

INTRODUCTION

Worldwide, the uses of herbal medicines are increasing rapidly among the community (Chen et al., 2010). Quality control of herbal medicines involves ensuring the correct identify of the plant intended as both consumers and manufacturers stand to lose from misidentification and contamination in terms of harms to health or legal responsibility and loss of future sales. Therefore, an identification tool is required to differentiate species of medical plants as each species has different medical functions as reported for Hibiscus in Maganha et al. (2009) and Malvaviscus (cited in Yeasmin et al., 2014).

Hibiscus, which has medicinal properties (Healthline Media, 2005; Kim, 2005; Maganha et al., 2009) belongs to the Malvaceae plant family which has 243 genera and 4225 species of plants. Another genus of this family is the Malvaviscus (Berry, 2017), and the species *Malvaviscus arboreus var drummondii* is reported to have antioxidant activity (Yeasmin et al., 2014). Thus, an efficient way of identification of this group of plants would be beneficial.

In past decades, recognition of plant species has traditionally been based on morphological characters (Duminil & Michele, 2009). However, morphological diagnostic approaches have several limitations; plants of the same species may show different morphologies caused by local adaptation and variability of the genes as well as phenotypic plasticity (Duminil & Michele, 2009). Hence, molecular methods started to become popular to distinguish species of plants (Duminil & Michele, 2009). Among these methods, DNA barcoding is one which provides an accurate and reliable authentication of species (Sundar, Nithaniyal, Raju Balaji & Madasamy, 2016).

DNA barcoding method aims to be a universal method which uses a short sequence of DNA (marker) from a specific portion of the genome of the plant in order to identify the species of plant (Herbert, Cywinska, Ball & deWaard, 2003). However, a single universal locus has not been obtained for identification of all plant species. Fazekas et al. (2012) suggested *rbcL* (Ribulose bisphosphate carboxylase large chain)

and *matK* (Megakaryocyte-associated tyrosine-protein kinase) as core DNA barcodes as well as *trnH-psbA* spacer (between the gene for tRNA histidine and the gene coding for photosystem II protein D1) and *ITS2* (Internal Transcribed Spacer) as supplementary regions. On the other hand Chen et al, (2010), suggested ITS2 as the standard DNA barcode to identify species of medicinal plants. But in Gonzalez et al, (2009), the *rbcL* marker and *trnH-psbA* marker showed the best performance as DNA barcodes for identification of tropical juvenile plants present in the Amazonian forest. Due to the different identification ability of the 4 markers in different plant groups, this experiment aims to find out which marker is efficient in distinguishing plant species of the family Malvaceae.

In this experiment, *matK* and *trnH-psbA* will be used to characterize and tested for their ability to identify *Hibiscus rosa sinensis* variants with a single layer of white petals, the *H. cannabinus L* as well as a member of the genus Malvaviviscus, i.*e. Malvaviscus arboreus var. drummondii* previously identified as *Hibiscus malvaviscus* (Uniprot, 2002).

CHAPTER 2

LITERATURE REVIEW

2.1 Medicinal Plants - *Malvaceae* Family

Nowadays, traditional medicine is widely used. 80% of the world's population uses it to cure disease. Plants are not only used as herbal medicine (Yeasmin et al., 2014), but can also be the raw material for production of pharmaceutical drugs (Chen et al., 2010). The rise in use of the herbal medicine is increasing the demand towards botanical raw materials (Parveen, Gafner, Techen, Murch, & Khan, 2016). *Hibiscus* and *Malvaviscus* are genera in the *Malvaceae* family (Berry, 2017) which have medicinal properties (Healthline Media, 2005; Kim, 2005; Maganha et al., 2009).

2.2 Introduction of Three Genera within *Malvaceae*

2.2.1 *Hibiscus rosa-sinensis*

The *Hibiscus rosa-sinensis*, a tropical hibiscus also known as China rose or Chinese hibiscus is a common type of ornamental plant which has different colored flowers such as red, white and pink. The *H. rosa-sinensis* has a funnel-like flower with a central stamen and can divided into those which have flowers with a single layer of petals or those with a double layer of petals. It also has ovate dark green leaves (Missouri Botanical Garden, n.d.; SFGATE, 2002). However, wild *H. rosa-sinensis* cannot be found because the *H. rosa-sinensis* is a hybrid species from hybridization of 8 or more different species which originate from India and Islands in the Pacific Ocean (Sparnaaij, 1973 ; SFGATE, 2002).

2.2.2 *Malvaviscus arboreus var drummondii*

The *Malvaviscus arboreus var drummondii* also called Turk's Cap or Texas mallow also belong to the *Malvaceae* family. They are evergreen and bloom the whole year round (Berry, 2017; Lady Bird Johnson Wildflower center, 2017; Kamper, 2015). It is a type of shrub which has bright red flowers with five petals that fold together and a stamen in the center of the flower. The leaves of the *M. arboreus var drummondii* are heart-shaped, palmately veined, small and the edges toothed (Kamper, 2015). The *M. arboreus var drummondii* has also been named as *Hibiscus malvaviscus* (Uniprot, 2002).

2.2.3 *Hibiscus cannabinus L*

The *Hibiscus cannabinus L* also called kenaf which belong to the *Malvaceae* family is an annual, fast growing plant which grow from woody to herbaceous (Duke, 1983). It is also included in the bast fibre group. It has a prickly stem reaching around 4.2m in height and their leaves are long-petiolate with five lance-shaped lobes which are toothed. The flower of *H. cannabinus L* has 5 sepals of white, yellow, or purple with purple centres, held on short stalks. It has medical use as a purgative, and in treating coughs, and stomach ache (Encyclopædia Britannica, 2013; Duke, 1983).

2.3 Authentication of Medical Herbs

The 21st century, has seen increased popularity of traditional medicines due to the realization of side effects of synthetic drugs (Prakash, Jyoti, Kumar, Kumar, Manna, 2013). According to Ganie et al. (2015), more than 1000 companies are producing herbal products with more than US$60 billion annual revenue. This indicates the high demand for traditional medicine. This may be a cause for the presence of adulterants or substitute plants as raw material to make medicine as there is a need to fulfill the demand (Parveen et al., 2016; Chen et al., 2010; Wijayasiriwardena, 2013), while resources maybe limited or expensive. Adulterants or substitute plants are plants which have similar appearance (Wijayasiriwardena, 2013). The use of adulteration material becomes a problem which will influence the quality and efficacy of the traditional medicine as well as the safety of the medicine (Ganie, Upadhyay, Das,

Sharma, 2015; Parveen et al., 2016). Therefore, powerful and accurate authentication methods are required to identify between the real medical plants and the substitute plants to ensure the drugs are of good quality and potency (Chen et al., 2010).

2.4 Traditional Approach--- Morphology-based authentication

There are different traditional approaches that can be used to identify the species of plants such as morphology-based method, macroscopic and microscopic method, and chemical analysis (Parveen et al., 2016; Wijayasiriwardena, 2013). The traditional approach that is often used for identification of plant species is morphology (Kress et al., 2005). It is based on the appearance of the plant such as the color of the flower, the shape of the leaves, and the height of the plant to differentiate the species of the plant (Hartvig, Czako, Kjær, Nielsen, & Theilade, 2015). Morphology-based method is useful in identification of the plants species and is a cost effective method (Wijayasiriwardena, 2013). However, there are still some limitations of this method. First, the variation of the genes due to hybridization between closely related plants will lead the plants to change their appearance. This will lead to wrong identification. Besides in some plants certain morphological characters like flowers may not always be present. So, it is difficult to differentiate the species of plant based on their morphology (Hebert, Cywinska, Ball, & deWaard, 2003). Identifying plant species using morphology-based methods also require expert knowledge that is not often available (Hartvig et al, 2015; Duminil & Michele, 2009). Additionally different genus of plants may be similar in appearance, which will then be identified wrongly into the same group (Hartvig et al, 2015; Duminil & Michele, 2009). Therefore, an alternative or additional method is required for identification.

2.5 Alternative Approach---DNA-based Method

Due to the limitation of the morphology-based method, alternative DNA-based methods are being more commonly used. These methods use the unique deoxyribonucleic acid (DNA) present in each plant to differentiate the plant species (Ganie et al, 2015). The DNA sources can be divided into 3 genome groups which are nuclear, chloroplast and mitochondrial to identify plant species. The mitochondrial genome is the one which is least used because it does not have sufficient polymorphic

information to use for plant species identification (Duminil & Michele, 2009). The most used is the nuclear genome (Duminil & Michele, 2009). The common methods to identify differences in the genome include RAPD (Random amplified polymorphic DNA), AFLP (Amplified fragment length polymorphism), microsatellite and DNA barcoding (Arif, & Khan, 2010). However, no one is the ideal method in identifying the plant species because different techniques are suitable to use in different situations (Arif, & Khan, 2009; Ganie et al, 2015).Table 2.5.1 shows the benefits and limitations of DNA barcoding to two other DNA based methods (Hartvig et al, 2015).

Table 2.5.1 Differences between the DNA-barcoding and other DNA-based Methods (adapted from Hartvig et al, 2015).

	RAPD	AFLP	DNA Barcoding
Genomic DNA required	15-30ng	200-300ng	30-50ng
Reproducibility	Very low	High	Very high
Applicability in plant authentication	Yes, but should not use (genetic diversity marker)	Yes, but should not use (a prefect genetic diversity marker)	Yes
Cloning/sequencing	No	No	Yes
Use of radioactivity	No	Yes	No
Detection of alleles	Generally no	Generally no	Yes
Cost	low	high	medium
Ease of use of automation	Very easy	Difficult	Easy

2.6 DNA Barcoding

DNA barcoding can be defined as a technique which uses a short and standard gene region to identify the plant species. The gene region that is used should be universally present in all plants (Ganie et al, 2015; Liu, Ci, Li, Li, Conran, & Li, 2017; Elansary, Ashfaq, Ali, & Yessoufou, 2017). It has been proposed as an important supplemental method toward the traditional morphology-based method. There are several advantages of using DNA barcoding to identify plant species. First, DNA barcoding does not require expert taxonomic knowledge. It is a method which refers to the database that has already been established to identify the specific species. Secondly, DNA barcoding can be carried out with only a small tissue sampled from the organism; it does not even require the reproductive part (flower, fruit), which is often