ASSESSING *matK* AND *ITS2* DNA BARCODES FOR IDENTIFYING *Hibiscus rosa-sinensis* MORPHOLOGICAL VARIANTS

SEOW YEE LOON

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

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

YEAR 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 proposal 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.

_____	_____
Name	Signature
_____	_____
I.D.Number	Date

DECLARATION FORM

I hereby declare that the work in this thesis is my own except for quotations and summaries which have been duly acknowledged, and completed under the supervision of Prof./Assoc. Prof./Dr./Mr./Ms. (and co-supervision of Prof./Assoc. Prof./Dr./Mr./Ms. if any)

SEOW YEE LOON (I13004032)

14TH JULY 2017

DR. GEETA SELVARAJAH

(CO-SUPERVISOR if any)

ABSTRACT

A major problem in conservation and ecological studies is the limited number of taxonomists available to identify the vast numbers of flora and fauna, amongst them the abundant and globally distributed Hibiscus species. In this study, the discriminant analysis of morphological traits of *Hibiscus* species also revealed the importance of floral characteristics in distinguishing between species in this genus as well as subspecies, and the difficulty of identification in the absence of flowers. Hence, DNA barcoding is a potential tool to overcome the burden of the taxonomists and provide a rapid detection to be used by non-taxonomists. *ITS2* and *matK were* tested for the identification of *Hibiscus rosa-sinensis* variants, together with *Hibiscus schizopetalus*. After plant DNA was extracted from leaves, the barcode loci was amplified, sequenced and the data analysed. Despite the good quality genomic DNA extracted, only *matK* PCR product had sequences of good quality. BLAST analysis yielded ambiguous identification perhaps due to requirement to reexamine the separation of genus in this taxon. The best fit nucleotide substitution model was T92 model instead of the widely used K2P model which was probably due to the unequal bases frequencies (A≠T≠G≠C). The phylogenetic trees were able to distinguish between *Hibiscus rosa-sinensis* and *Hibiscus schizopetalus*, but not the morphological variants. However, the data complemented morphological identification in the absence of flowers indicating that molecular data including other loci could contribute to identification and thus studies and management of species diversity.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AIC	*Akaike Information Criterion*
AFLP	Amplified fragment length polymorphism
BIC	*Bayesian Information Criterion*
BLAST	Basic Local Alignment Search Tool
bp	base pair
CBOL	Consortium for Barcode of Life
*CO*1	Cytochrome c I oxidase
DNA	Deoxyribonucleic acid
EDTA	Ethylene Diamine Tetraacetic Acid
i.e.	In effect
ITS	*Internal transcribed spacer*
K2P	*Kimura 2 parameters*
matK	*Maturase K*
MEGA 5	Molecular Evolutionary Genetics Analysis 5
MEGA 6	Molecular Evolutionary Genetics Analysis 6
M.W.	Molecular weight
NaCl	Sodium chloride
NaOH	Sodium hydroxide
n.d.	No date
NJ	Neighbour joining
PCR	Polymerase chain reaction
rbcL	*ribulose-bisphosphate carboxylase*
RAPD	Random amplified polymorphic DNA
RFLP	Restriction fragment length polymorphism
SDS	Sodium Dodecyl Sulfate
UV	Ultraviolet

CHAPTER 1

INTRODUCTION

"To waste, to destroy our natural resources, to skin and exhaust the land instead of using it so as to increase its usefulness, will result in undermining in the days of our children the very prosperity which we ought by right to hand down to them amplified and developed."- Theodore Roosevelt

Thus it is important that we study and conserve the biodiversity of our environment. It has been the norm that conservation and ecological studies, examine wildlife and plants under a pristine environment. However, as the world is changing and there is increased urbanization, it is also important to examine the biodiversity in urban environments, which can be another avenue of conservation.

One genus commonly found in the urban landscape is the Hibiscus including the species *Hibiscus rosa-sinensis.* It belongs to the *Malvaceae* family which consists of 88 genera and 2300 species across the world (Ayanbamiji, Ogundipe, & Olowokudejo, 2012). The genus Hibiscus exhibits considerable taxonomic complexity and is so heterogeneous that it is hard to identify distinguishing features between some species (Fryxell, 1997; Alam, Pasha, & Ahmad, 2006). Furthermore within a single species such as *Hibiscus rosa-sinensis,* are plants with morphologically different forms, for example flowers which are single, semi double to fully double and from white to red (Bates, 1961).

Identification thus is challenging, yet it is crucial in order to determine the rarity and to proceed with the conservation measurements. Traditionally, expert taxonomists will provide their services in species classification and identification whereby the information will be used as references that aid in all biological studies (Muhammad Qaiser, Khali, Shinwari, 2015). However, identifying every taxon of organisms is a difficult task and it may be subjective depending on the taxonomist (Muhammad Qaiser et al., 2015) and there is also the taxonomic impediment (De Carvalho et al., 2007), i.e. a lack of taxonomist.

To overcome the problem, molecular methods should be used to compliment or support morphological methods. One such method is DNA barcoding – a quick and reliable analysis technique which allows the identification of species by using a short, standardized genome fragment and compare to a library of known sequences (Muhammad Qaiser et al., 2015; John, Thomas, Edward, Melchias, & Prabhu, 2011; Kress, Garcia-Robledo, Uriarte, & Erickson, 2015). The DNA barcoding approach enables species identification for non-taxonomist and is a crucial tool for addressing fundamental questions in the field of ecology, evolution and conservation studies (Kress et al., 2015; Hebert & Gregory, 2005).

Even though DNA barcoding is now a popular and effective tool in plant species identification, there is still no consensus on the plant DNA barcoding loci (Li et al., 2014). As choosing the appropriate gene for the purpose of barcoding is difficult (Cowan & Fay, 2012; Hollingsworth, Graham, & Little, 2011), few candidates were recommended – *trnH-psbA* by Kress, Wurdack, Zimmer, Weigt, & Janzen (2005); *second internal transcribed spacer, ITS2* by Chen et al. (2010); *ribulose-1,5-biphosphate carboxylase oxygenase large subunit (rbcL)* and *maturase K (matK)* by Li et al. (2011), Bafeel et al. (2011), and the CBOL Plant Working Group (2009).

In this study, we aim to examine the ability of *maturase K* (*matK*) and the *second internal transcribed spacer* (*ITS2*) to differentiate the morphological variants of *Hibiscus rosa-sinensis* using barcoding gap and tree topology phylogenetic methods to strive in conservation and ecological studies.

CHAPTER 2

LITERATURE REVIEW

2.1 PLANT CONSERVATION

The planet's functionality and humans' survival solely depends on plants. Without plants, there will be no life. Recent plant conservation activities aim to achieve a situation whereby human activities are able to co-exist with plant life diversity, in which plants in return will be able to improve our environment's well-being (Convention on Biological Diversity, n.d.; Bureau of Land Management, 2016). Through the collection of data and studies of plants' environment and diversity, we can achieve the final goal of plant conservation whether it is for the benefit of the present or future generation or the benefit of environment stability (Bureau of Land Management, 2016; Department of Botany, 2017).

2.2 IDENTIFY SPECIES USING DNA BARCODING

Identifying species and assigning them to their respective taxon has been a fundamental and crucial process throughout history (Kaplan, 2001; Haider, 2011; Pereira, Carneiro, & Amorim, 2008). The traditional and conventional method for species identification is based on their morphological characteristics (Haider, 2011; Rossatto, Casanova, Kolb, & Bruno, 2011) which hinge on the knowledge, experience and interpretation of expert taxonomists in order to provide accurate identification (Muhammad Qaiser, Ali, & Zabta, 2015; Rossatto et al., 2011). In addition, certain species are closely related and share certain similarities in morphological characteristics whereas some may appear similar but are different taxonomic entities (Pereira et al., 2008; Duminil & Michele, 2009). This has complicated the morphological-based approach (Muhammad Qaiser et al., 2011; Pereira et al., 2008). To overcome the problem, the strategy of using a molecular approach developed; simplifying and supporting the data of the morphological approach (Haider, 2011; Pereira et al., 2008).

Molecular approaches such as restriction fragment length polymorphisms (RFLP), amplified fragment length polymorphisms (AFLP), random amplified polymorphic DNA (RAPD) or polymerase chain reaction (PCR) kick-started the era of species identification based on DNA analysis (Duminil & Michele, 2009; Pereira et al., 2008). These molecular based approaches also provided significant information on intraspecific genotypic variation (Pereira et al., 2008; Arif et al., 2010). However, the occurrence of migration, random genetic drift or reproduction will influence the genetic variation within species. Not to mention the differing rate of mutation and recombination within each locus (Pereira et al., 2008).

2.3 SELECTION OF DNA BARCODES IN PLANTS

A short segment of standardized region DNA being utilized as a tool of species identification is known as DNA barcodes (Cowan, Chase, Kress, & Savolainen, 2006; Hollingsworth, Graham, & Little, 2011; Kress, Wurdack, Zimmer, Weigt, & Janzen, 2005; CBOL Plant Working Group, 2009). The purpose of DNA barcoding is to create a feasible and inexpensive method of species identification for non-taxonomists and create resources of DNA sequence library among shared communities which will be used for organismal identification and taxonomic clarification (Hollingsworth et al., 2011; Muhammad Qaiser et al., 2011).

The ideal DNA barcode will be a short fragment of sequence enough for sequencing and easy to operate (Cowan et al., 2006; Hollingsworth et al., 2011). Subsequently, the DNA region from the short fragment must have certain variability to be distinguish among species and less intraspecific (Cowan et al., 2006).

The approach of using the short mitochondrial gene *chytochrome C oxidase subunit I* (*COI*) as the standard DNA barcode for animal species identification was a huge success (Hollingsworth et al., 2011; CBOL Plat Working Group, 2009; Kress et al., 2005; Braukmann, Kuzmina, Sills, Zakharov, & Hebert, 2017) due to the characteristics of protein-coding region and the display of high discriminating power (Hollingsworth et al., 2011). However, *COI* gene is not suitable for plant species identification as plant

mitochondrial DNA had displayed a low rate of nucleotide substitution which limited *COI* gene (Cowan et al., 2006; Hollingsworth et al., 2011; CBOL Plant Working Group, 2009; Braukmann et al., 2017). Morover, plant with high level of mitochondrial content displayed frequent rearrangement among DNA sequences, from gene to genome (Cowan et al., 2006). Therefore, alternative approach must be figured in order to aid in plant species identification.

Several DNA barcodes were proposed as the standard DNA barcode for plant species identification. Plastid *trnH-psbA* spacer is a good barcoding candidate due to its short sequence length which is widely accepted in a variety of plants and easy to be amplified (Cowan et al., 2006; Kress et al., 2005) However, *trnH-psbA* has a potential flaw which leads to difficulties in the alignment due to short sequences (Cowan et al., 2006) and the possibility of unable to differentiate plant with low levels of genetic difference among species due to the present of high intraspecific polymorphisms (Pang et al., 2012; Whitlock, Hale, & Groff, 2010). For *maturase K* (*matK*), was proposed by Heckenhauer, Barfuss and Samuel (2016) as it had high reliability and high success rate of primers amplification. However, a low amplification and sequencing success of *matK* has been reported (Kress et al., 2010; Hollingsworth et al., 2011).

The internal transcribed spacer 2 (*ITS2*) has also been proposed as the candidate of barcoding as it displayed an ease in amplification and sequencing and variability for plant species identification (Kress et al., 2005; Yao et al., 2010; Han et al., 2013). The presence of its secondary structure information is also crucial for species identification (Han et al., 2013).

As up till now, there are DNA barcodes for plant species, however, there is also a possibility that two or more combination of DNA barcodes are required in order to achieve a higher level of species identification (Kress et al., 2005; CBOL Plant Working Group, 2009).

2.4 SPECIES DISCRIMINATION ANALYSIS

Various methods such as Basic Local Alignment, barcode gap analysis, tree topology using distance or characters and genealogical methods (Cummings et al., 2008) are available for taxon discrimination and species identification purposes. Similarity-based methods such as Basic Local Alignment Tool (BLAST) (Altschul et al., 1997), assign query barcodes to species based on the amount of DNA bases within the sequences have in common. The distance based barcode gap is founded on degree of DNA sequence variation within and between species and assigns a cutoff value (the 'barcode gap') to divide taxonomic units (Hebert, Stoeckle, Zemlak, & Francis, 2004; Ball, Hebert, Burian, & Webb, 2005; Zhou, Adamowicz, Jacobus, DeWalt, & Hebert, 2009). Other distance methods may be monophyly-based, which is the recovery of species depends on discrete clades (monophyly) on a phylogenetic tree (Hebert, Ratnasingham, & deWaard, 2003). Tree based methods could also be character based on characters such as parsimony where the phylogenetic tree that minimizes the total number of character-state changes is to be preferred (Fitch, 1971). Another character based method, CAOS treats the nucleotide sites in a DNA sequence as characters which are diagnostic (Sarkar, Planet, Bael, Stanley, Siddall, DeSalle, & Figurski, 2002).

2.5 CHALLENGES OF DNA BARCODING

Challenges to DNA barcoding plants range from DNA extraction, identifying useful barcoding loci to the method of assigning a plant sample to a species based on sequence data. Malvaceae such as hibiscus contain secondary compounds including polysaccharides (Echevarría-Machado, Sanchez-Cach, Hernandez-Zepeda, Rivera-Madrid, & Moreno-Valenzuela, 2005), which could inhibit DNA extraction and amplification (Moyo et al.,2008). Despite the existence of recommended DNA barcodes (CBOL Plant Working Group, 2009), lack of universality induces researchers to continue testing loci for plant species identification (Cowan & Fay, 2012) and building a complete reference plant biodiversity database (Cowan & Fay, 2012). Research persists to overcome the barriers to DNA barcoding, pushed by the possible uses of DNA barcoding.