

**DNA BARCODING ANGIOSPERMS:
IDENTIFYING PORTULACA SPECIES USING *MATK* AND *ITS2***

YU HAI PING

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

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

2018

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.

YU HAI PING

Name

I15007867

I.D.Number



Signature

6/8/2018

Date

ACKNOWLEDGEMENT

I would first like to thank my thesis advisor Assoc. Prof. Dr. Geeta Selvarajah. The door to Dr Geeta's office was always open whenever I ran into a trouble spot or had a question about my research. She consistently allowed this paper to be my own work, but steered me in the right the direction whenever she thought I needed it.

I would also like to thank Dr Tham Yew Joon and Ms. Emily Quek as the second reader of this thesis, and I am gratefully indebted to their very valuable comments on this thesis.

Lastly, I want to thank my fellow colleagues especially Hing, Zien for lending me their hands whenever I faced problems. Besides, I am very grateful to have companions from all colleagues during lab works which gave me memorable days.

ABSTRACT

Weeds are invading agricultural farms causing impacts to both industry and consumers, not only in terms of economic loss through crops reduction and increase of labour costs in elimination of weeds but also in their invasive characteristic which destroy the native biodiversity in the farm. DNA Barcoding is a fast and accurate species identification system which utilizes standardize short DNA sequence as an universal barcode known as marker. Earlier stage of invasive weeds identification can be done with the ease by DNA Barcoding and thus serve to inform weed management for effective solution. DNA barcoding while being a widely used method, has had problems in its use such as lack of consensus on gene loci. In this study, *Portulaca oleracea* (both ornamental and wildtype) and *Portulaca grandiflora* were sampled for DNA Barcoding with the use of *matK* and *ITS2* loci respectively whereby the efficiency of the gene loci were evaluated based on assessments of sequence quality, and levels of species discrimination. DNA of all three replicates of each of the *Portulaca* species were successfully extracted in triplicates using the Edwards method. Both *matK* and *ITS2* were successfully amplified and sequenced followed by NCBI BLAST. The BLAST results showed identification to species level while the unexpected hit of ornamental *P. oleracea* (*ITS2* amplicon) to *P. umbraticola* revealed the limitation of DNA barcoding which is incomplete reference sequence from public database. The phylogeny tree implies capability of both *matK* and *ITS2* in discriminating *Portulaca* species as unique clades are formed for each taxon.

TABLE OF CONTENT

	Page
NON-PLAGIARISM DECLARATION	ii
DECLARATION	iii
ACKNOWLEDGEMENT	vi
ABSTRACT	v
TABLE OF CONTENT	vi
LIST OF TABLES	viii
LIST OF FIGURES	ix
LIST OF ABBREVIATIONS	x
CHAPTER	
1 INTRODUCTION	1
2 LITERATURE REVIEW	3
2.1 WHY ARE WEEDS A PROBLEM?	3
2.2 INTRODUCTION TO RESEARCH SPECIES	3
2.2.1 <i>Portulaca oleracea</i>	3
2.2.2 <i>Portulaca grandiflora</i>	4
2.3 THE IMPORTANCE OF TAXONOMY	5
2.4 TRADITIONAL APPROACH	6
2.5 ADVANCED APPROACH – DNA BARCODING	7
2.5.1 DNA Barcoding for integrated weed management	7
2.6 DNA PRIMERS FOR PCR (BARCODING REGION)	8
2.6.1 Chloroplast <i>maturase K</i> gene (<i>matK</i>)	8
2.6.2 Internal transcribed spacer 2 (<i>ITS2</i>)	9
3 MATERIALS AND METHODS	10
3.1 SAMPLE COLLECTION	10
3.2 DNA EXTRACTION	10
3.2.1 DNA quality assessment by agarose gel electrophoresis	11

3.3	PCR PREPARATION	12
3.3.1	Preparation of agarose gel	14
	electrophoresis to assess the quality of pcr product	
3.4	DATA ANALYSIS	14
3.4.1	Sequence editing	14
3.4.2	BLAST analysis	14
3.4.3	ITS Database	14
3.5	MULTIPLE SEQUENCE ALIGNMENT	15
3.6	PHYLOGENY TREE CONSTRUCTION	15
4	RESULTS ANALYSIS	
4.1	DNA EXTRACTION ASESMENT	16
4.2	PCR PRODUCTS ASESMENT	17
4.3	SEQUENCE ANALYSIS	19
4.4	BLAST analysis	20
4.5	ITS Database	22
4.6	MULTIPLE SEQUENCE ALIGNMENT	23
4.7	PHYLOGENY TREE	24- 27
5	DISCUSSION	28
5.1	ANALYSIS OF DNA EXTRACTION	28
5.2	ANALYSIS OF PCR PRODUCTS	28
5.3	SEQUENCE QUALITY	29
5.4	BLAST ANALYSIS	29
5.5	PHYLOGNY TREE	31
5.6	LIMITATIONS AND RECOMMENDATIONS	31
6	CONCLUSION	33
	REFERENCES	33
	APPENDICES	38- 50

LIST OF TABLES

Tables		Page
3.1	Primer Sequence of <i>matK</i> and <i>ITS2</i> used for amplification	12
3.2	Reagents present in the PCR mix used for PCR reaction	13
3.3	PCR amplification conditions for <i>matK</i> and <i>ITS2</i> marker	13
4.1	BLAST results for <i>matK</i> sequences of all three species	21
4.2	BLAST results for <i>ITS2</i> sequences of all three species.	22
4.3	Structure prediction using <i>ITS</i> Database (http://its2.bioapps.biozentrum.uni-wuerzburg.de/). Only the best hit is selected. GI (model) is the homology model of query sequence	23

LIST OF FIGURES

Figures	Page
2.1	4
2.2	4
4.1	16
4.2	17
4.3	18
4.4	19
4.5	20
4.6	23
4.7	24
4.8	25
4.9	26
4.10	26
4.11	27
4.12	27

LIST OF ABBREVIATIONS

°C	Celsius
(v/v)	Concentrations
Bp	Base pairs
Cm	Centimetres
CO1	Cytochrome c oxidase subunit I
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
<i>ITS2</i>	Internal Transcribed Spacer
g	grams
Kb	Kilobyte
<i>matK</i>	Megakaryocyte-associated tyrosine-protein kinase
mL	Millilitre
min	Minutes
MSA	Multiple Sequence Alignment
NaCl	Sodium Chloride
PCR	Polymerase Chain Reaction
<i>rbcL</i>	Ribulose biphosphate carboxylase
RNA	Ribonucleic acid
Rpm	Revolutions per minute
RT	Room temperature
SDS	Sodium lauryl sulphate
SYBR	cyanine dye
TAE	mixture of Tris base, acetic acid and EDTA.
Tris- HCl	Tris(hydroxymethyl) aminomethane hydrochloride
µl	Microliter

CHAPTER 1

INTRODUCTION

For the last 50 years, the world population growth has multiplied at a rapid rate and is believed to continue to grow to reach 9 billion by 2050 (Bureau, n.d.). A growing population means a growing global demand for food and thus it is necessary to construct a better agricultural system to meet the future food demand.

Portulaca is one of the genus of *Portulacaceae* family, comprising up to 115 species (The Plant List, 2010). Some *Portulaca* species are ornamental like the *Portulaca grandiflora* and *Portulaca pilosa* which are commonly known as moss flower and “kiss me quick” respectively (Dyer, n.d.). On the other hand, the species *Portulaca oleracea* known as Purslane is a crop-invading weed which can infest field crop and cost undesired impacts (Uddin et al., 2014). It is a prolific seeder which can spread across wide area with the ability to protect its seed outwards from mother plant, rather than just releasing them. It's also a strong species whereby any piece of the plant left behind will re-root and continue to grow (Purslane, *Portulaca Oleracea*, Research & Weed Control Advice). Therefore, it requires urgent attention from agriculturalists to come out with an effective solution in dealing with the damages that *P. oleracea* will bring.

Weeds invade agricultural farms and cause impacts to both industry and consumers, not only in terms of economic loss through crops reduction and labour costs in elimination of weeds but also in their invasive characteristic which destroy the native biodiversity in the farm (Weeds in Australia, n.d.). Weeds cannot be easily identified in a timely manner by traditional approach alone based on morphological characters and may sometimes cause confusion with crops that are not weeds (Queensland Government, 2016). Therefore, an identification tool is required to differentiate between weeds and farm plants at molecular level for effective treatment in terms of shorter time and lower cost to eradicate them as weeds are strongly adaptive species which may naturalize and hybridize with the closely related locals causing gene pool interruption within the crops.

DNA barcoding is by now the most widely method used by taxonomist as it's a fast and accurate species identification system (CBOL Plant Working Group et al., 2009a). This method utilizes standardize short DNA sequence as universal barcode known as marker. The marker is different for various targets like *cytochrome c oxidase 1* (CO1) for animals while *matK* and *rbcL* for plants (Sarvananda, 2018). However, there has been no agreement on which regions(s) should be used as a universal marker for barcoding land plants (CBOL Plant Working Group et al., 2009a).

In this experiment, *matK* (Megakaryocyte-associated tyrosine-protein kinase) and *ITS2* (Internal Transcribed Spacer) were evaluated and tested for their ability to identify and differentiate *Portulaca oleracea* (both ornamental and wildtype) and *Portulaca grandiflora* using the DNA Barcoding approach based on assessments of sequence quality, and levels of species discrimination (CBOL Plant Working Group et al., 2009b).

CHAPTER 2

LITERATURE REVIEW

2.1 PROBLEM OF WEEDS

No plant is a “weed” in nature. They’re naturally strong growers found in nature that help to protect soil from erosion as well as to replenish organic matter in restoring soil fertility (Mark, 2010). However, human activities create the existence of “weed” by categorizing weeds as wild plants that outgrow and compete with cultivated plants and have undesirable impacts which outweigh their advantages (Dwight ,n.d.).

There are many ways how weeds affect plantations. Weeds generally reduce crop yield by competing with crops for nutrients, light, water as well as space (National Institute Of Open School, 2013). Its vigorous growth habit is the reason for significant crop damage. The need to carry out weed control increases the cost of cultivation at which 30 percent of the total expenditure for crop production is on tillage operations (Manitoba, n.d.) as more labour is required. Besides, weeds encourage insects’ population as it provides shelter and food for insects during harsh conditions. What is worse is that certain weeds may produce chemical substances which are toxic to crop plants (Dwight, n.d.).

2.2 INTRODUCTION TO RESEARCH SPECIES

2.2.1 *Portulaca oleracea*

The *Portulaca oleracea* (purslane) which is an indigenous species belonging to the Southeast Asia region particularly in Persia and India (Sandra, n.d.). It is now widely spread throughout the world as a weed while at the same time being eaten extensively as food and medicine around the Mediterranean and tropical Asian countries (Zhou et al., 2015). The *P. oleracea* is characterised by its succulent leaves and stems with yellow flowers. The succulent properties of the purslane allow it to be drought resistant and it literally can grow just anywhere from fertile soil to poor soil.



Figure 2.1 *Portulaca oleracea* (purslane); shoot, with blooms (CABI, n.d.).

2.2.2 *Portulaca grandiflora*

The *Portulaca grandiflora* has a flatter leaf than the nearly cylindrical leaf of the *P. oleracea* (Gilman & Howe, 1999). It is often cultivated ornamentally because of its attractive flowers which have variety of bright colours including yellow, orange, white, pinkish purple and white (Plantinfo, n.d.).



Figure2.2 *Portulaca grandiflora*, photo by Daryl Mitchell, CC-BY-SA-2.0 (NC STATE University, n.d.).

2.3 THE IMPORTANCE OF TAXONOMY

To date, there are about 1.5 million species being identified and described formally in scientific literature (University of Chicago, 2017). There are approximately 2 billion living species on Earth estimated by group of researchers from University of Arizona in a new published paper, *The Quarterly Review of Biology* (September 2017). This number of estimated living species is a thousand-fold of the current number of described species which tell us that current progress of species identification needs to be improved. The job of identifying all living species is one of the most important basics in science.

Global biodiversity is being lost at a rate which is far from optimistic as a result of irrational human activities. A basic understanding on the diversity and how many species there are on the planet is absolutely fundamental as it serves as knowledge for us to maintain the ecosystem that humanity depends upon (Sweetlove, 2011). By understanding the components of biodiversity, conservation and sustainable use of ecosystem can be done effectively by appropriate decision-making based on comprehensive establishment of action on areas to be protected. Additionally, taxonomic information is essential for agencies and border authorities to detect invasive alien species from native species for control of the consequences that can be caused by them (Sample, 2004).

2.4 TRADITIONAL APPROACH

Taxonomy has generally been executed using discontinuities in analogue (i.e., graded morphological) traits to infer species boundaries, an approach that has generated a total of 1.7 million taxonomic hypotheses over 250 years (Hebert & Gregory, 2005). Several different traditional approaches have been used for species identification such as morphology-based method, macroscopic and microscopic method, and chemical analysis (DeSalle, Egan, Siddall, 2005). Identification of plant species is often based on morphology of appearance of plant such as the leaf segmentation, colour of the flower, structure of stems (Hartvig, Czako, Kjær, Nielsen, & Theilade, 2015). It's a cost effective method as morphology can be readily observe without handling or harming the organisms (McGowan, n.d.).

However, there are many downfalls of this method where one of the issues with morphological systematics is that the similarity of phenotypic characteristics among some groups of organisms can cause confusion. There are generally no universal methods agreed by taxonomists on how to define one morphological feature differentiating one species from another (Friedheim-Sophie, 2016). The lack of a conventional agreement on morphological definition and species classification creates a barrier to a better taxonomy system (Friedheim-Sophie, 2016).

2.5 ADVANCED APPROACH – DNA BARCODING

DNA Barcoding is a novel approach in species identification with the use of specific short DNA sequences which has been proposed and initiated to facilitate the taxonomy system (Kress, Wurdack, Zimmer, Weigt, & Janzen, 2005). DNA Barcoding approaches to taxon diagnosis exploit diversity among DNA sequences to identify organisms (Hebert, Cywinska, Ball) and these sequences are viewed as “barcodes” that are present in every cell. The stages of the process are (1) sampling of the species, (2) DNA extraction, (3) PCR amplification of DNA barcode using specific markers, (4) sequencing of DNA barcode (5) BLAST analysis (6) phylogeny construction.

DNA Barcoding has been proposed as an important supplemental method toward the traditional morphology-based method, to identify plant species in this case. There are several advantages of using this approach to identify plant species. Firstly, DNA Barcoding does not require expert taxonomists to do the work as one can collect the sample from targeted species and send to a barcoding laboratory for species identification. This allows farmers to better understand what species are invading their farm and implement better weed management strategies. Secondly, this method is applicable for all life stages where invasive weeds can be identified before it grows into seed producing stage. Thirdly, it can differentiate plant species which cannot be identified based on morphology.

While DNA barcoding has many benefits, it is not without shortcomings. Among these is that a single barcoding locus has not been found to work across all species (CBOL Plant Working Group, 2009). Similar statements have also been reported in a paper “DNA Barcoding: Promise and Pitfalls” that whether “one gene fits all” is open to debate as there has been no agreement on which region(s) should be used as a universal marker for barcoding plants