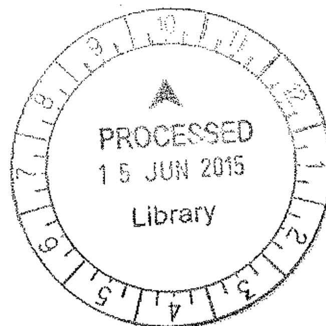
DST 001200
TAN SRI ABDUL MAJID LIBRARY

DNA BARCODING *JASMINUM SAMBAC, JASMINUM NITIDUM, AND NYCTANTHES ARBOR-TRISTIS* BY USING ITS2 AND *trnH-psbA* GENES

CHANG KEK NING

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



TP
248
2
CHA
2015

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

2015

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 School of Applied Science 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.

Chang Kek Ning	
Name	Signature
I12001777	8th MAY 2015
I.D.Number	Date

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to my research supervisor, Dr. Geeta Selvarajah for her guidance and support of this research project. Even with the many problems and difficulties during this project, she was still patient and provided useful information in this project. I feel very grateful to her for spending so much of time to review my works. With her assistance, I was able to complete this project within 14 weeks. I also want to thank our Head of Programme, Mr. Lim Chia Hong for giving me this opportunity to do a final year project. This enables me to apply knowledge that I learnt from this course and practice research. Last but not least, I want to thank the laboratory assistants for their help in providing the apparatus and materials that I needed in this project.

ABSTRACT

Jasminum and *Nyctanthes* are from Oleaceae family which are widely cultivated and distributed all around the world especially Asian countries. The chemical constituents from the plants of Oleaceae family *Jasminum* and *Nyctanthes* have high commercial value because it can be used in the pharmaceutical industry. *Jasminum* and *Nyctanthes* are a source of essential oil and perfume due to its fragrant smell. They can also be processed to make tea and food flavorings. *Jasminum* and *Nyctanthes* though from the same family are in different genus though the phylogenetic position of *Nyctanthes* has been debated. By using DNA barcoding, we can determine the differences within and between species. Three different species of plants which are *Jasminum sambac*, *Jasminum nitidum* and *Nyctanthes arbor-tristis* were analysed using the Internal Transcribed Spacer 2 (ITS2) and *trnH-psbA*. The DNA samples were amplified using Polymerase Chain Reaction (PCR) for ITS2 and *trnH-psbA* gene sequences. The PCR products were sequenced, and then the results were analysed using T-Coffee, MEGA 6 and BLAST. Based on the results, *J.sambac* and *J.nitidum* shows higher similarity (0.128 for ITS2; 0.032 for *trnH-psbA*) compared to *N.arbor-tristis* and *Jasminum* (Average=0.301 for ITS2; Average=0.237 for *trnH-psbA*) because *J.sambac* and *J.nitidum* come from the same genus, and this is one of the criteria for a good barcode.

TABLE OF CONTENT

	PAGE
NON-PLAGIARISM DECLARATION	**i**
ACKNOWLEDGEMENTS	**ii**
ABSTRACT	**iii**
TABLE OF CONTENT	**iv**
LIST OF TABLES	**vi**
LIST OF FIGURES	**vii**
LIST OF ABBREAVIATIONS	**viii**
CHAPTER	
1. **INTRODUCTION**	**1**
2. **LITERATURE REVIEW**	**3**
2.1 DNA barcode	3
2.1.1 DNA barcode for plants	4
2.2 Molecular evolution	5
2.2.1 Nucleotide substitution models	6
2.3 Oleaceae family	7
2.3.1 Genus *Jasminum*	7
2.3.2 Genus *Nyctanthes*	8
3. **MATERIALS AND METHODS/METHODOLOGY**	**9**
3.1 Samples collection and identification	9
3.2	
3.2.1 Isolation of DNA	9
3.2.2 Analysis of genomic DNA samples using gel electrophoresis	10
3.3	
3.3.1 Amplifying DNA by Polymerase Chain Reaction	11
3.3.2 Analysis of PCR products using gel electrophoresis	12
3.4 Sequencing and data analysis	12

4. **RESULT** **14**
4.1 Analysis of genomic DNA samples 14
4.2 Analysis of PCR products 15
4.3 Sequencing and data analysis 16
4.3.1 Nucleotide substitution models 17
4.3.2 Nucleotide composition 17
4.3.3 DNA distance matrix 18
4.3.4 Barcoding gap 19
4.3.5 Phylogenetic tree 19
4.4 BLAST analysis 20

5. **DISCUSSION** **22**
5.1 Analysis of genomic DNA samples 22
5.2 Analysis of PCR products 22
5.3 Data analysis using MEGA 6 23
5.3.1 Nucleotide composition 23
5.3.2 DNA distance matrix and barcoding gap 24
5.3.3 Phylogenetic tree 25
5.4 BLAST analysis 26
5.5 Recommendations 26

6. **CONCLUSION** **27**

REFERENCES **28**

APPENDICES **34**

LIST OF TABLES

Table		Page
2.1	General nucleotide substitution models and some criteria	6
4.1	The best-fitting model was selected in subsequent analyses of the data set	17
4.2	The average nucleotide composition of *J.sambac, J.nitidum* and *N.arbor-tristis*	18
4.3	The distance matrix of ITS2 gene for all samples	19
4.4	The distance matrix of *trnH-psbA* gene for all samples	20
4.5	The barcoding gap between two different species was calculated by using distance matrix of ITS2 and *trnH-psbA*	21
4.6	The BLAST results after inserted the query separately based on ITS2 gene for all samples	23
4.7	The BLAST results after inserted the query separately based on *trnH-psbA* gene for all samples	23

LIST OF FIGURES

Figure		Page
2.1	General process of DNA barcoding	4
4.1	The banding pattern for the genomic DNA samples from three different species	14
4.2	The banding pattern of PCR products for ITS2 gene amplification from three different species	15
4.3	The banding pattern of PCR products for *trnH-psbA* gene amplification from three different species	16
4.4	The phylogenetic tree was constructed based on ITS2 using the Neighbor-Joining method	21
4.5	The phylogenetic tree was constructed based on *trnH-psbA* using the Neighbor-Joining method	22

LIST OF ABBREVIATIONS

BLAST	Basic Local Alignment Search Tool
CO1	cytochrome c oxidase 1 gene
dH_2O	Distilled water
DNA	Deoxyribonucleic acid
EDTA	Ethylene Diamine Tetraacetic Acid
ITS2	Internal Transcribe Spacer 2
matK	maturase K
NaCl	Sodium chloride
PCR	Polymerase chain reaction
rbcL	RuBisCo large subunit
RNA	Ribonucleic acid
SDS	Sodium dodecyl sulphate
TBE buffer	Tris/Borate/EDTA buffer
TE buffer	Tris/EDTA buffer
TER buffer	Tris/EDTA buffer with RNase A
°C	degrees Celsius
g	gram
µL	microliter
mL	millilitre
Min	minute
nm	nanometre
S	second
V	Volt

CHAPTER 1 INTRODUCTION

Before development of DNA barcoding, taxonomist usually used morphological characters (Science Learn, 2009) such as size, shape, color and other parts of body to classify living organisms. Taxonomy characterizes organisms that have similar features according the binomial nomenclature to identify the genus and species of each organism. However, the change of global climate and habitat destruction is causing extinction of species (Bellard, Bertelsmeier, Leadley, Thuiller & Courchamp, 2012) even before taxonomists are able to identify much of the biodiversity. Hence, DNA barcoding is being developed as a quick way to identify the species based on the unique and short DNA sequence (Cold Spring Harbor Laboratory, 2014). In addition, DNA barcoding is able to investigate the adulteration in meat and plant products, market substitution in seafood, and the factors that cause infectious disease (Galimberti et al., 2012).

The mitochondria *cytochrome c oxidase 1* gene served as a standard barcode to differentiate many animal groups (Bellard, Bertelsmeier, Leadley, Thuiller & Courchamp, 2012). However, this gene is not suitable to identify plant species because the substitution rate of this gene is too slow (Wagner & Ulrich-Merzenich, 2013). For plant species, RuBisCo large subunit (*rbcL*) and maturase K (*matK*) from chloroplast are usually used to identify plant species (Fiser & Buzan, 2014). The advantage of using *rcbL* to identify plant species is the presence of this gene in most land plants and the ability to easily amplify it (Jawdat, Elias, Al-Faoury, & Al-Safadi, 2013). But the mutation rate of this gene is low (Wu, et al, 2010) so it will be difficult to differentiate between species. The *matK* loci too can be easily amplified and the discrimination capability is higher than *rbcL* (Wu & et al, 2010). Although *rcbL* and *matK* have been used widely, they have not always been successful in discriminating species (Jeanson, Labat & Little, 2011), and many other loci have been suggested to be more efficient as a DNA barcode such as internal transcribed spacer 2 (ITS2) and *trnH-psbA* (Gao, Liu, Yao, Song, & Chen, n. d.).

Three different species of plant which are *J.sambac*, *J.nitidum* and *N.arbor-tristis* will be analysed by using the ITS2 and *trnH-psbA*. All three of these species are from the Oleaceae family, but *J.sambac* and *J.nitidum* are in a different genus from *N.arbor-tristis*. ITS2 from the nuclear ribosomal DNA region and *trnH-psbA* from the chloroplast, both have high discrimination capability at species level (Techen, Parveen, Pan, & Khan, 2014). Moreover, the identification rate of *trnH-psbA*+ITS2 is higher than *matK*+ *rbcL* to identify most of the land plant species (Pang et al., 2012).

The aims of this study are:

1. To obtain the ITS2 and *trnH-psbA* sequences from three different species which are *Jasminum sambac, Jasminum nitidum* and *Nyctanthes arbor-tristis*.
2. To compare ITS2 and *trnH-psbA* sequences between and within these three different species of sample plants.

CHAPTER 2 LITERATURE REVIEW

2.1 DNA BARCODE

Hebert with his research group first used DNA barcoding in 2003 and published a research paper, titled "Biological identifications through DNA barcodes" (Hebert, Cywinska, Ball & DeWaard, 2003). It was a new concept which used a short and standardized region from a genomic DNA to classify every species. This unique identification region from DNA is similar with the Universal Product Code (UPC) that can be used to identify, which product you purchase in a supermarket.

In the past, taxonomists faced difficulties in using morphology to classify living organisms (Scamardella, 1999) especially those specimens which did not have the structures required for classification. Also, the phenotypic plasticity of species may lead to errors occurring during identification (Pires & Marinoni, 2010). Every species has their growth process and their appearances will also keep changing due to genetic and environmental factors. So, the reliability of morphological method to classify organisms is very low. The lack of expertise also is one of the main issues of traditional taxonomy (Pires & Marinoni, 2010). Hence, the advent of DNA barcoding can effectively identify various species based on the short DNA markers because every organism is made up of cells which contain DNA and DNA is quite stable. Another significant advantage brought by DNA barcoding is its ability to discriminate morphologically identical species.

There are two main functions of DNA barcode which is to discover novel genotypes that might form a foundation to assist subsequent species discovery and identifying the recognized species (Erickson et al., 2008). Although this method is quite similar to genomic sequencing, which also uses DNA sequencing, there are some differences which exist in their procedure. DNA barcoding is a highly standardize method, using a single set of primer for conserved gene which is present in all taxa (Makarova, 2014). Besides that, DNA barcoding usually involves sequencing a short fragment with an average of 600 base pairs (Makarova, 2014) instead of sequencing the whole genome of species. Thus, DNA barcoding became

more popular in classifying species because it is affordable, fast, and simple (Pires & Marinoni, 2010).The outline of DNA barcoding process is shown as the picture below:

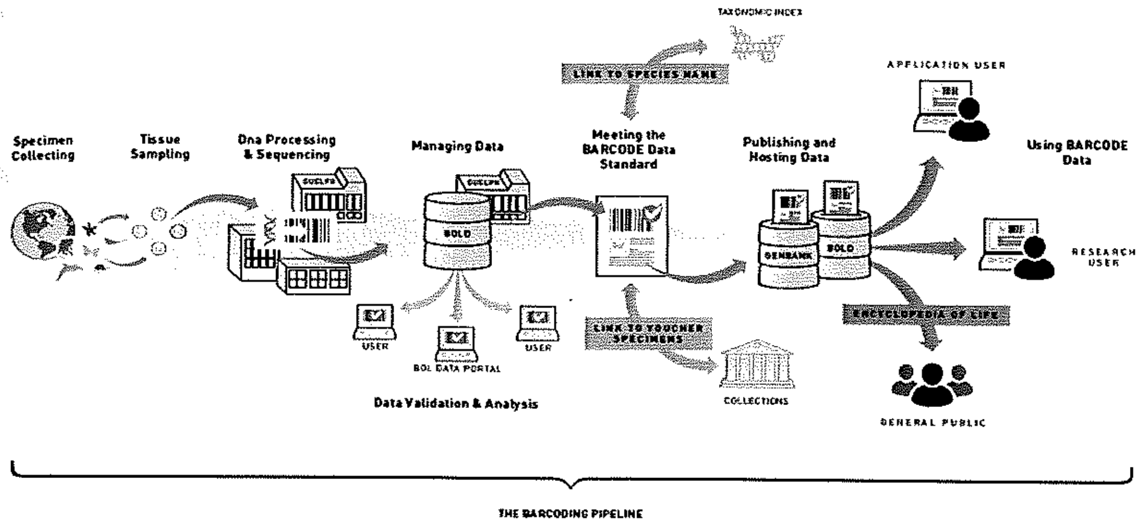


Figure 2.1 General processes of DNA barcoding.
(Consortium for the Barcode of Life, n. d.)

2.1.1 DNA Barcode for Plants

For plants, the mitochondrial DNA has a low substitution rate making it unsuitable to use *CO1* as a barcode region. There are many barcodes which can be used but scientist normally use *matK, rbcL, trnH-psbA* from plastid and also ITS from nuclear ribosomal (Kress & Erickson, 2012). However, there has been no restriction to use whichever DNA marker/gene to identify the plant species. A reliable barcode region should able to be retrieved a single primer pair and can be sequenced bidirectionally. The most important thing is the marker/gene should give high discrimination rate among species (Hollingsworth et al., 2009). The advantage of *rbcL* is the improvements done to the primer, making it easily retrieved among plant species. Besides that, *rbcL* is able to provide high-quality of bidirectional sequences after sequencing (Hollingsworth et al., 2009). The discrimination rate of *rbcL* is lower compared to ITS2 and *trnH-psbA* (Jeanson, Labat & Little, 2011). *matK* and *trnH-psbA* can usually provide high resolution of sequence data. However, the primer universality of *matK* is poorer compared to *rbcL* and *trnH-psbA*, and sometimes cannot provide a consistent bidirectional sequence (Hollingsworth et al., 2009). For ITS2, the advantages are high discrimination rate, easy to be amplified and

availability of conserved region to design universal primers (Yao et al., 2010). However, ITS2 has a lower resolving power compared to ITS among the close related species (Han et al., 2013). By compared the advantage and disadvantage for each DNA barcode, the DNA barcode that have been chosen for this study are ITS2 and also *trnH-psbA*. Many researches are using this combination because of the overall effectiveness and also the high discrimination rate that the barcodes give (Liu, Chen, Song, Zhang & Chen, 2012). ITS2 was selected to be one of the barcode markers because ITS2 is widely used in phylogenetic reconstruction at intraspecies or interspecies level (Enan, Fawri, Al-Deeb & Amiri, 2012).

2.2 MOLECULAR EVOLUTION

Molecular evolution is a process of changing genetic material over a long period of time which is normally caused by mutation at the gene level (Speed, Kechris & McWeeney, n. d.). Even though mutation can occur in every living organism at a high rate, it wouldn't be obvious in extant populations because of natural selection.

After studying molecular evolution, scientists are able to reconstruct the history of evolution based on their similarity level of biological molecules. Substitution mutation could lead to evolution at the molecular level. The information of the mutated strain is stored in the DNA of that particular individual or organism which will then be inherited by the next generation. Mutation at the molecular level could also lead to differences between many different individuals of the same species. This is where DNA barcoding comes in. DNA barcoding can easily identify the multiple unknown individual's phylogenetic relationship. Scientists are using DNA barcoding because they can process many samples at the same time and at a faster rate compared to genomic sequencing.

2.2.1 Nucleotide Substitution Models

Nucleotide substitution is a type of mutation which has a nucleotide base, changed to another base. Substitution can be divided by two types which is transition and transversion. Transition is the interchange of purine into purine (A⇔G) or pyrimidine into pyrimidine (C⇔T); transversion is the interchange between purine into pyrimidine and vice versa. Due to the substitution mutation, scientists developed models to find out the model which best fits the data that we are able to obtain. The main purpose of all these nucleotide substitution models is to computational analysing the evolutionary relationship based on nucleotide sequences alignment. There are many models can be tested and find out which model is the best fit for interpret a particular data set (Posada, 2012). A good model can make the reconstruction of evolutionary history more accurate. All these nucleotide substitution models can be used to calculate the probability change of nucleotide sequence. The table below show the general nucleotide substitution models and some certain criteria:

Table 2.1 General nucleotide substitution models and some criteria.

Model	Criteria
Jukes and Cantor 1969 (JC69) (Liò & Goldman, 1998)	All substitution have same probability with equal base frequencies
Kimura 2-parameter 1980 (K80) (Liò & Goldman, 1998)	The rate of transition and transversion are same with equal base frequencies
Felsenstein 84 1984 (F84) (Adam et al., 2014)	All substitution have same probability with variable base frequencies
Hasegawa Kishino Yano 1985 (HKY) (Adam et al., 2014)	The rate of transition and transversion are same with variable base frequencies
Tamura 1992 (T92) (Liò & Goldman, 1998)	Generalised the Kimura method by relaxing the assumption of equal base frequencies. This model applies in strong transition and transversion and GC content bias.
Tamura and Nei 1993 (TN93) (Liò & Goldman, 1998)	Distinct rate of transition and transversion with variable base frequencies