BIODIVERSITY AND MEDICINAL METABOLITES OF RARE AND ENDANGERED VIETNAMESE PLANTS USING MOLECULAR GENETICS AND PROTEOMICS

By

NGUYEN VAN HUY

Submitted in fulfilment for the degree of Doctor of Philosophy (PhD)
School of Life Sciences
Faculty of Science
University of Technology Sydney (UTS) Australia

February 2019

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, NGUYEN VAN HUY, declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Life Sciences, Faculty of Science, at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Production Note:
Signature removed prior to publication.

Nguyen Van Huy

15th February 2019

ACKNOWLEDGEMENT

I wish to thank my supervisors Dr. Luigi De Filippis and Dr. Fraser Torpy for their advice and assistance during the research program and thesis preparation period. Their tireless enthusiasm and encouragement have been an important contribution to my thesis. I would also like to thank the School of Life Sciences and the Graduate Research School, University of Technology Sydney for the opportunity and support afforded to me during the candidature.

I sincerely thank the following academic staff Dr. Matthew Padula, Gemma Armstrong, and Kun Xiao from the School of Life Science laboratories for their skill training and useful advice to solve technical problems. It is also a great honour to mention here the expert advice and consultations with Dr. Pham Thanh Loan from Faculty of Agro-forestry of Hung Vuong University.

I also highly appreciate the encouragement and support from many staff at the Hung Vuong University in general, and the staff of the Department of Scientific Research Planning and Technology in particular, who have taken over my jobs during my PhD candidature. I would also like to acknowledge and thank my colleagues and students; the full list of names cannot be listed here, but a particular thanks to Divya Vinod, Trieu Hoa, Nguyen Quang Trung, Cao Chi Cong, Le Van Quang, Doan Thanh Son, Bui Quang Tien, and Nguyen Quang for their time and insights during data collection and experimental implementation.

The Joint-scholarship between the Vietnam Ministry of Education and Training and the University of Technology Sydney is, of course, the most important part of being able to achieve this PhD. I indeed have to testify that I could not afford to do this research degree without their financial support.

Last but not least, I wish to express my great appreciation to my family for their patience, understanding and encouragement throughout my studies. Special thanks to my parents, especially my parents-in-law who have taken care of my children while I was studying in Australia. Finally, my heartfelt thanks go to my wife Yen Hoang for her constant support and encouragement throughout my PhD, without whom none of this work would have been possible.

Any remaining errors or shortcomings in this thesis are of course my responsibility.

TABLE OF CONTENT

ACKNOWLEDGEMENT ii
LIST OF FIGURES vii
LIST OF TABLES ix
ABSTRACT x
CHAPTER 1 1
INTRODUCTION 1
1.1 Overview 1
  1.1.1 Plant material 1
  1.1.2 Genetic diversity 3
  1.1.3 Metabolites 4
  1.1.4 Proteomics 5
1.2 Summary 7
1.3 Objective and scope of the study 7
1.4 Principal questions 9
1.5 Significance of the study 9
  1.5.1 Scientific 9
  1.5.2 Practical 9
CHAPTER 2 11
LITERATURE REVIEW 11
2.1 Plants and conservation 11
  2.1.1 General 11
  2.1.2 Medicinal plant conservation 12
  2.1.3 Concerns to medicinal plant conservation 13
  2.1.4 Strategy for medicinal plant species conservation 14
  2.1.5 Conclusion 15
2.2 Genetic diversity 16
  2.2.1 General 16
  2.2.2 Methods of measuring genetic diversity 18
  2.2.3 Analysis of genetic diversity from molecular data 22
2.3 Anti-oxidant enzymes and metabolites 24
  2.3.1 General 24
  2.3.2 Anti-oxidant reaction 25
  2.3.3 Formation of proteases 28
  2.3.4 Proline 30
  2.3.5 Glutathione 30
  2.3.6 Tocopherol 31
  2.3.7 Flavonoids 32
  2.3.8 Phospholipids and lipid peroxidation 34
  2.3.9 Malondialdehyde 35
  2.3.10 Conclusion 36
2.4 Proteomics 36

2.4.1 General ... 36
2.4.2 Plant proteomics ... 38
2.4.3 Medicinal plant proteomics ... 42
2.4.4 Proteomic strategy and approaches in secondary metabolism ... 46
2.4.5 Sample preparation, protein identification and bioinformatics ... 47
2.4.6 Conclusion ... 56

CHAPTER 3 ... 57

STUDY SITES AND SPECIES ... 57

3.1 An overview of *Cycas fugax* ... 57
3.1.1 Introduction ... 57
3.1.2 Taxonomy ... 57
3.1.3 Distribution and conservation status ... 58
3.1.4 Morphological characteristics ... 58

3.2 An overview of *Celastrus hindsii* research ... 59
3.2.1 Introduction ... 59
3.2.2 Taxonomy and morphology ... 61
3.2.3 Biological activity study ... 63
3.2.4 Breeding study and cultivation state in Vietnam ... 65

3.3 The study sites ... 65
3.3.1 Ba Vi National Park, Ha Noi City ... 67
3.3.2 Cat Ba National Park, Hai Phong City ... 69
3.3.3 Phu Tho province ... 70
3.3.4 Hoa Binh province ... 71

3.4 Summary and outlook ... 72

CHAPTER 4 ... 73

GENETIC DIVERSITY ... 73

4.1 Introduction ... 73

4.2 Material and methods ... 75
4.2.1 Study site ... 75
4.2.2 Sampling strategy ... 75
4.2.3 Sample collection and storage ... 76
4.2.4 DNA extraction ... 77
4.2.5 DNA qualification ... 78
4.2.6 Polymerase Chain Reaction ... 78
4.2.7 PCR preparation ... 79
4.2.8 Primer selection and optimisation of PCR ... 79
4.2.9 Gel electrophoresis ... 79
4.2.10 Band scoring ... 80
4.2.11 Statistic analysis ... 80

4.3 Results ... 81
4.3.1 DNA extraction ... 81
4.3.2 PCR optimisation ... 83
4.3.3 Intra-population genetic diversity ... 85
4.3.4 Inter – population genetic diversity ... 88

4.4 Discussion ... 92
4.4.1 DNA extraction ... 92
4.4.2 PCR optimisation ... 95

4.4.3 Intra-population genetic diversity 97
4.4.4 Inter-population genetic diversity 99

4.5 Conclusion **102**

CHAPTER 5 ......**103**

ENZYMES AND METABOLITES **103**

5.1 Introduction **103**

5.2 Materials and methods **105**
5.2.1 Leaf material 105
5.2.2 Protein content 106
5.2.3 Proteases 106
5.2.4 Proline 107
5.2.5 Glutathione (GHS) 107
5.2.6 Tocopherol 108
5.2.7 Flavonoids 109
5.2.8 Phospholipids 109
5.2.9 Malondialdehyde (MDA) 110
5.2.10 Replications and statistics 110

5.3 Results **111**
5.3.1 Protein content 111
5.3.2 Protease 111
5.3.3 Proline and glutathione 112
5.3.4 Tocopherol and flavonoid 113
5.3.5 Phospholipid and malonyldialdehyde (MDA) 113

5.4 Discussion **114**
5.4.1 Protein content 114
5.4.2 Protease 115
5.4.3 Proline and glutathione 117
5.4.4 Tocopherol and flavonoids 119
5.4.5 Phospholipids and malondialdehyde (MDA) 123

5.5 Conclusion **124**

CHAPTER 6 ......**126**

PROTEOMICS **126**

6.1 Introduction **126**

6.2 Material and method **129**
6.2.1 Plant material 129
6.2.2 Chemicals 129
6.2.3 Protein extraction 129
6.2.4 Quantification of proteins 130
6.2.5 Protein fractionation and detection 131
6.2.6 Protein identification and bioinformatics 131

6.3 Results **133**
6.3.1 Protein extraction 133
6.3.2. Proteomic profiling 135
6.3.3 Analyses of differential expression protein 140

6.4 Discussion **140**
6.4.1 Protein identification 140

6.4.2 Proteins facilitating growth and development through amino acid and nuclease metabolism 143
6.4.3 Proteins facilitating growth and development through stress and defence response 150
6.4.4 Other growth and development-related proteins 153
6.4.5 Proteins involving secondary metabolite biosynthetic and post-translational modification 155

6.5 Conclusion 159

CHAPTER 7 160

DISCUSSION AND CONCLUSIONS 160

7.1 Conservation genetics 160
7.1.1 Methodological consideration 160
7.1.2 Research limitations 161
7.1.3 Genetic diversity of a rare and endangered plant – C. fugax 162
7.1.4 Conservation genetics of a medicinal plant – C. hindsii 163

7.2 Anti-oxidative metabolites as key traits for medicinal plants 164

7.3 Proteomics 167
7.3.1. Regulatory proteins in biosynthesis of secondary metabolites 167
7.3.2. Production of valuable secondary metabolites 168

7.4 Potential for new bioactive compounds 170
7.4.1 Proteins with antitumor / anti-inflammatory activity and therapeutic applications 172
7.4.2. Proteins overexpressed in tumour cells and their therapeutic applications 173

7.5 Plant propagation, breeding, and cultivation 174
7.5.1. Propagation and re-introduction of C. fugax 174
7.5.2 Plant breeding for superior plants in C. hindsii 175
7.5.3 Cultivation and sustainable product production 176

7.6 Conclusions and future research perspective 178
7.6.1 Genomic study 178
7.6.2. Metabolic study 179
7.6.3 Proteomic study 179
7.6.4 Management and sustainability 180
7.6.5 Bioprocessing and molecular pharming 181
7.6.6 Next-generation sequencing (NGS) 182
7.6.7 Future research not often contemplated 182

REFERENCES 184

APPENDICES 237

LIST OF FIGURES

Figure 1.1 Nucleic Acids and Proteins, and their roles in studying cellular processes....6

Figure 2.1 Overview of the reaction leading to the formation of ROS..........................26

Figure 2.2 Classes of natural antioxidants... 27

Figure 2.3 Simplified schematic of the flavonoid pathway (*Source*: Pourcel et al.........33

Figure 2.4 Workflow of a proteomic experiment...45

Figure 2.5 The preparation of protein extracts from plant tissues 49

Figure 3.1 Vegetative and reproductive organs of *C. fugax* 59

Figure 3.2 Subgeneric classification of *Celastrus* and distribution and nearest living relative of *C. hindsii* in leaf morphology shown by cluster analysis (between-groups linkage)...62

Figure 3.3 Vegetative and reproductive organs of *C. hindsii* 63

Figure 3.4 Study locations of *C. fugax* (only in Phu Tho province) and *C. hindsii* (in four provinces in the north of Vietnam...66

Figure 4.1 Example of a RAPD-PCR and RAMP=PCR profile among *C. fugax* and *C. hindsii* individuals 84

Figure 4.2 Multi dimensional scaling (MDS) - (Figure A) and Principle component analysis (PCA)– (Figure B) of *C. fugax* using RAPD-PCR data and RAMP-PCR data. 90

Figure 4.3 Dendrograms (cluster analysis using Nei genetic diversity) in RAPD-PCR (A) and RAMP-PCR (B) of the rare and endangered plant *C. fugax*.. 91

Figure 4.4 Multi-dimensional scaling (MDS) – (Figure A) and Principle component analysis (PCA) - (Figure B) of *C. hindsii* using RAPD-PCR data and RAMP-PCR data.. 91

Figure 4.5 Dendrograms (cluster analysis using Nei genetic diversity) in RAPD-PCR (A) and RAMP-PCR (B) of the medicinal plant *C. hindsii*. 92

Figure 5.1 Percentage water and leaf thickness (A), protein content on the basis of air dry weight (ADW), dry wieght (DW) and leaf area (LA) (B) in narrow leaf (NL) and broad leaf (BL)...111

Figure 5.2 Acid protease (A) and neutral protease (B) activity on a protein basis and a dry weight basis in NL and BL extracts...112

Figure 5.3 Proline (A) and glutathione (B) content on a protein basis and a ADW basis in NL and BL extracts..112

Figure 5.4 Tocopherol (A) and flavonoid (B) content on a protein basis and a dry weight basis in NL and BL extracts...113

Figure 5.5 Phospholipid (A) and MDA (B) content on a protein basis and a dry weight basis in NL and BL extracts. 114

Figure 6.1 Protein profile: 134

Figure 6.2 Protein distribution and profiling of *C. hindsii* expressed in NL (Figure B) and BL (Figure A) phenotypes. 135

Figure 6.3 Functional classification of differentially expressed proteins from NL and BL of *C. hindsii*. Proteins were categorised according to their gene ontology (GO) using Blast2Go software 138

Figure 6.4 Top hit species distribution based on the number of BLAST top hits 139

LIST OF TABLES

Table 2.1 Active compounds and botanical oils of some medicinal plants in Vietnam . 13

Table 2.2 Basic statistical concepts on genomic data for genetic diversity assessment . 23

Table 2.3 Identified proteins and their role in secondary metabolism in medicinal plants using a proteomics approach 43

Table 2.4 Characteristics of some 2DE rehydration buffers used in protein solubilization 52

Table 3.1 Origin and description of samples collected from four populations of *C. fugax* in Tram Than Commune for RAPD-PCR and RAMP-PCR. 66

Table 3.2 Origin and description of samples collected from four populations of *C. hindsii* in four provinces for RAPD-PCR and RAMP-PCR. 67

Table 4.1 Quality and quantity of DNA extracted from young and mature leaves of *C. hindsii*. 82

Table 4.2 Quality and quantity of DNA extracted from mature leaves of *C. fugax* and young leaves of *C. hindsii*. 82

Table 4.3 Optimization of PCR parameters 83

Table 4.4 DNA polymorphism between four populations of *C. fugax* detected by RAPD-PCR. 86

Table 4.5 DNA polymorphism between four populations of *C. fugax* detected by RAMP-PCR. 87

Table 4.6 DNA polymorphism between four populations of *C. hindsii* detected by RAPD-PCR. 87

Table 4.7 DNA polymorphism between four populations of *C. hindsii* detected by RAMP-PCR. 88

Table 4.8 DNA polymorphism and population statistics and Mantel test. 89

Table 5.1 Morphological parameters to classify NL and BL of *C. hindsii* 105

Table 5.2 Total flavonoid content in some medicinal plants 121

Table 6.1 Protein yield from different plant species extracted by BPP 134

Table 6.2 Identification of differentially expressed proteins 137

Table 6.3 Proteins classified as high score and low score based on Mascot and BLAST search. 142

Table 7.1 Summary of proteins detected by this thesis with pharmaceutical and human therapeutic implications 171

ABSTRACT

This thesis reports on an investigation into the molecular biology and metabolism of two species: *Cycas fugax*, a critically endangered cycad; and *Celastrus hindsii*, a traditional medicinal plant. The genetic diversity of both species was investigated to understand distributions, important for conservation and prioritisation of efforts in both species. Investigations on phytochemical accumulation and biosynthesis of metabolites and enzymes associated with the two types of phenotypes of *C. hindsii* (broad leaf-BL and narrow leaf-NL) was undertaken to identify the medicinally superior sources of material for possible commercial utilisation.

The genetic diversity of *C. fugax* and *C. hindsii* was determined using Randomly Amplified Polymorphic DNA (RAPD) and Randomly Amplified Microsatellite Polymorphism (RAMP) techniques. The latter molecular method proved more useful in distinguishing between individuals and populations. Recommendations on future conservation and management for the two species are identified.

The project then concentrated on *C. hindsii.* The concentration of important anti-oxidative metabolites was investigated and compared between NL and BL phenotypes using spectrophotometric methods. BL had a significantly higher concentration of α-tocopherol, flavonoid, phospholipid and possibly glutathione (based on an air-dry weight (ADW) and protein basis). Stress biomarkers such as proline and malondialdehyde (MDA) were elevated in NL, indicating their roles in response to stress. Total protein increase in BL correlated with a reduction of proteases, particularly acid protease. In contrast, the enzyme and metabolite increase in NL may be associated with tissue degradation and an indication of stress in the natural forest area.

Differentially expressed proteins were analysed between NL and BL to understand the mechanisms which facilitate the growth and development of BL, particularly the accumulation of some potentially valuable phytochemicals. By using two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) coupled with mass spectrometer (MS), a range of upregulated proteins was identified. Some proteins were indicative of growth promotion and development in BL plants via diverse biological processes and molecular mechanisms, such as flavonoid biosynthesis, stress response, disease resistance and nucleic acid metabolism. The proteins identified supported plant growth and development processes and are of pharmaceutical importance, with possible therapeutic applications in mammalian disease treatment, particularly in cancer therapy.

To conclude, *C. fugax* urgently requires a genetic rescue plan to protect it from extinction. Genetic conservation for *C. hindsiii* is also necessary to promote breeding, cultivation and system development for possible medicines. BL of *C. hindsii,* which contains more diverse and superior concentrations of important phytochemicals and pharmaceutically related proteins should be positioned as an important commercial source for natural drug development.