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

LIST OF FIGURES

LIST OF TABLES

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.

CHAPTER 1

INTRODUCTION

1.1 Overview

1.1.1 Plant material

Cycas fugax is one of the rarest species in the genus *Cycas.* The family Cycadaceae are a primitive group of vascular plants (Hendricks 1987; Hill, Nguyen & Loc 2004; Cousins & Witkowski 2017). *C. fugax* is distinguished from other members of the genus by very few long leaves (1-3 leaves), widely spaced leaflets with very narrow tapering bases, small male cones with microsporophylls that have a terminal spine, and megasporophylls with elongated teeth (Hill et al. 2004).

Cycad species have been used for centuries by diverse cultural groups around the world, mostly as a source of starch during droughts and famines. In some regions, cycads are used for medico-magical practices, gum, fibre production, and are traded for traditional medicine, despite the toxicity of seeds. The toxins need to be eliminated before consumption (Cousins & Witkowski 2017; Cousins et al. 2011). Generally, there has been an increase in ethnobotanical studies of cycads in Central America, India, and South Africa (Krishnamurthy et al. 2013; Bamigboye Samuel et al. 2017; Shiba et al. 2016; Pérez-Farrera et al. 2006). The traditional use of cycads in Vietnam is however poorly known. This is particularly so for rare and endangered cycad species in need of conservation.

C. fugax is classified as critically endangered (CR) (possibly extinct) in the 2010 IUCN Red List (Osborne et al. 2010) with the species being distributed in a very restricted area in Phu Tho Northern Province, Vietnam. The species was discovered in the late 1990's, and its populations have declined dramatically and are near extinction as a consequence of natural habitat destruction, mainly by agricultural encroachment (Osborne et al. 2010). The only remaining populations are extremely fragmented in the wild or in home gardens in villages near their original habitat. This species is urgently in need evaluation of its conservation status. When populations are larger the possibility of exploitation of its pharmaceutical properties could be explored. At any rate, strong protective measures to save it from extinction are required.

Celastrus hindsii Benth belongs to the genus *Celastrus* Linnaeus and the family Celastraceae. The genus is a deciduous or evergreen woody vine, featured by alternate simple leaves, axillary or terminal cymes with mostly pale green bisexual or unisexual flowers, 3- to 6-seeded capsules and seeds covered by a distinct red or orange fleshy aril (Mu et al. 2012). Due to variation in leaf and flower morphology, some species in this genus have been defined without consistent agreement in the literature. Hence, Liang et al. (2016) suggested that molecular phylogenetic investigations are required to understand the relationships within the genus, and confirm existing subgeneric classifications.

According to Mu, Zhao & Zhang (2012), *C. hindsii* exhibits a close morphological similarity with *C. monospermus* and *C. tonkinensis* but it can still be distinguished by its broadly elliptic (vs narrowly elliptic) leaves and transverse (vs. smooth) valves (Muoi et al. 2009; Zhixiang & Funston 2008). *C. hindsii* has recently been cultivated throughout Vietnam, particularly in the Northern provinces (Muoi et al. 2009). *C. hindsii* crops have been established to harvest raw leaf material for medicinal purposes and can earn roughly AUD 2,000 per year/hectare for the first three years. The estimated income for the following years would be higher as increased productivity and is likely to be achieved (Loan 2015). Due to their morphological variation, particularly in leaf blade shape, a selective collecting method of local people for folk medicine purposes has been used. Two phenotypic variants have been recognised and distinguished; a narrow leaf phenotype (NL) which is considered less important, and a broad leaf phenotype (BL) which is preferred and recommended for traditional medicinal practices and trade in the local markets.

The plants of *Celastrus* are well-known for their medicinal properties, and it has been used for treatments of diseases such as fever, chills, joint pain, edema, rheumatoid arthritis, and bacterial infection. Their pharmaceutical potential is attributed to diverse secondary metabolites, such as sesquiterpenes, alkaloids, triterpenes, diterpenes, and flavonoids (Zhang et al. 2009). Of those species, *C. hindsii* is intensively used in folk medicine in Vietnam, mainly by the Muong people in Northern Provinces to treat ulcers, tumours and inflammation (Muoi et al. 2009; Zhixiang & Funston 2008). The pharmaceutical importance of *C. hindsii* have been supported by scientific evidence which showed that this species contains a variety of bioactive compounds such as vitamin C, vitamin E, sesquiterpenes, triterpenes, alkaloids and flavonoids (Yao-Haur Kuo, Chen & Kuo 1995; Ly, Shimoyamada & Yamauchi 2006; Hu et al. 2014). These secondary metabolites have been studied to develop modern disease therapeutics (Pan et al. 2013) and have importance in plant stress physiology, particularly in plant defence systems against biotic and abiotic stresses (Edreva et al. 2008).

C. hindsii has been cultivated in geographically separated populations under diverse conditions of climate, soil, fertilisers, etc. However, the wild-type (probably NL phenotype) is rarely found now, perhaps due to overexploitation by local people for domestic use and trade. *C. hindsii* broad leaf type has been preferred for collection by local people for folk medicines. If recognised phenotypes are proven to be significantly different in metabolite levels, this could be due to genetic or environmental factors (Shulaev et al. 2008; Rodziewicz et al. 2014). Morphological traits often have disadvantages for distinguishing genetic variants, and molecular studies might provide insights into phenotypic variation and assist the selection of populations for conservation and utilisation (Martínez-Esteso et al. 2015). Recent studies on *C. hindsii* have examined the chemical structure and bioactivity of secondary metabolites extracted from randomly collected samples. However, this type of information may be insufficient to assist a conservation and breeding program which usually need integrated research on related molecular biological aspects of the target species (Abril et al. 2011; Ahmad et al. 2016).

1.1.2 Genetic diversity

Genetic diversity represents the heritable variations distributed within and among populations of a species, and is regulated by evolutionary processes such as mutation, selection, genetic drift and the species mating system (Govindaraj et al. 2015). Although genetic diversity is encoded in the DNA of an organism, it is often not an accurate measure if the survival of the wild population is directly affected (Chevin & Hoffmann 2017). Unlikely, genetic diversity in a cultivated population, plants are susceptible to anthropogenic selection, breeding systems, and other crop-related conditions. As a result, it is vital to maintain the fitness of the wild populations in high genetic diversity level may keep them away from the risk of extinction and to provide a healthy plant system for cultivation (Govindaraj et al. 2015).

Genetic diversity can be assessed at intraspecies and interspecies levels by morphological, biochemical and molecular techniques. The first two methods might incur higher cost, time and resources, and fail to detect variation (Mondini et al. 2009). DNA-based techniques analyze variations directly at the DNA level, eliminating most environmental influences (Pathak & Abido 2014; Porth & El-Kassaby 2014). DNA-based techniques also only require a small amount of plant material. Molecular techniques use classes of DNA-based markers applied to genetic diversity assessment, but are also used for cultivar fingerprinting, and phylogenetic analysis (Collard & Mackill 2008; Currò et al. 2010).

These DNA-based markers can be grouped into three categories based on the method of detection; hybridisation-based, polymerase chain reaction (PCR)-based, and DNA sequence-based. The earliest of molecular techniques, namely Restriction Fragment Length Polymorphism (RFLP), detected genetic variation based on restriction-hybridisation techniques with the use of restriction enzymes that recognised and cut specific short sequences of DNA (Mondini et al. 2009). This method, although often revealing high polymorphism, is restrictive due to high cost, complications in interpretation and is time consuming (Bardakci 2001). In contrast, PCR-based markers have become widely used and have led to the development of several types of popular genetic analyses such as Random Amplified Polymorphic DNA (RAPD) and Random Amplified Microsatellite Polymorphism (RAMP).

The first PCR-based molecular markers were RAPD and RAMP. RAPD-PCR examines genetic variation through random amplification by short primes (10 base long synthetic oligonucleotides), RAMP-PCR uses combinations of microsatellite and RAPD primers (Welsh & Mcclelland 1990; Williams et al. 1990). There are no requirements for cloning and sequencing, and these two methods have gained importance due to their simplicity and efficiency. However, RAMP-PCR has proved more advantageous as it probably creates more specific data than does RAPD-PCR due to the longer SSR based primers, which enable higher stringency amplification, therefore reduce problems with reproducibility, a common criticism of low-stringency RAPD analyses (Yang et al. 1996). On the other hand, RAPD markers have some advantages of being employed as a method

of distinguishing the genetic relationships across genera and species of organisms (Al-Rawashdeh 2011).

Due to their high sensitivity, RAPD-PCR and RAMP-PCR are suitable for various applications in plant biology, particularly genetic diversity analysis to conserve and select suitable populations for utilisation. Using RAPD and ISSR primers, Harish et al. (2014) estimated the genetic diversity within and among different populations of an endangered medicinal plant *Commiphora wightii*; Kabir et al. (2014) assessed genetic relatedness among 20 elite accessions, and Adhikari, Bandopadhyay & Ghosh (2013) investigated the polymorphism among ten elite Indian cultivars of *Cymbopogon* (aromatic grass of essential oil trade types). These markers were also used to distinguish the wild and cultivated accessions (Khanna et al. 2014), and measured genetic distance between populations (Al-Rawashdeh 2011; Kazemi & Hajizadeh 2012). Similarly, RAMP has been broadly applied to discover the genetic diversity and genetic relationship of popular crops such as rice (Hoang et al. 2009), pomegranate (Zhao et al. 2013a), capsicum (Rai et al. 2013), almond and apricot progenies (Rasouli & Karimi 2015). Data extracted from genetic studies of *Clinacanthus nutans* (Ismail et al. 2016), *Litchi chinensis* (Long et al. 2015), *Lonicera japonica* (Fu et al. 2013) and *Moringa oleifera* (Avila-Treviño et al. 2017) have been supportive of agricultural practices and conservation strategies of these medicinal plants.

1.1.3 Metabolites

Genetic diversity information potentially provides conservationists and plant breeders with options to develop, through selection and breeding, new and more productive crops, that are resistant to biotic and abiotic stresses (Baruah et al. 2017). Maintaining a healthy defence system is essential for preventing oxidative stresses which are mostly eliminated by antioxidants such as vitamin C, vitamin E, and phenolic compounds (Georgieva et al. 2017; Noctor, Lelarge-Trouverie & Mhamdi 2015). Although many studies have focused on characterising the antioxidant activity of the medicinal plant *C. hindsii*, its intraspecific variation has not been conducted yet. Molecular markers may be useful for tolerance and cultivar selection in conservation and breeding programs, as suggested by Arbona et al. (2013), Rodziewicz et al. (2014), and Nakabayashi & Saito (2015).

The term metabolomics has been defined as the identification and quantitation of all low molecular weight metabolites in a given organism, tissue or cell type (Arbona et al. 2013; Nakabayashi & Saito 2015). Till now, there has not been a comprehensive laboratory approach to discover all compounds of interest but must rely on an appropriate combination of extraction and detection techniques (Noctor et al. 2015). Metabolite profiling is the most popular technique but can discover only a fraction of the metabolome. However, it allows evaluation of many samples in parallel, and to test effects on specific cellular processes (Shulaev et al. 2008).

A comprehensive coverage of the enormous range of metabolites present in an organism can be achieved when employing several analytical techniques, that often consist of a separation technique (e.g. high performance liquid chromatography (HPLC)) coupled to a detection device (e.g. mass spectrometer (MS)) (Yang et al. 2009; Nakabayashi & Saito 2015). The application of these techniques depends on the research question, as they require both time and labour, and are costly (Moniruzzaman et al. 2014). Plants are still a challenge since no comprehensive database exists for secondary metabolite analysis, particularly for medicinal species. As a result, a high throughput screening method (targeted profiling) which caters for a large number of samples in duplicate may be suitable for investigation of a target metabolite and describe intraspecific diversity and geographic patterns of variation (Shulaev et al. 2008). Stratil, Klejdus & Kuban (2006) reported that using spectrophotometric measurements with specific analytical reagents, at present, is adequate to obtain sufficient information for evaluation of contents of secondary metabolites in a comparative study. Although the profile of some anti-oxidative metabolites extracted from different locations has been reported in separate studies (Yao-Haur Kuo, Chen & Kuo 1995; Ly, Shimoyamada & Yamauchi 2006; Hu et al. 2014), intraspecific secondary metabolite variation in *C. hindsii* has not been studied.

1.1.4 Proteomics

The presence of secondary metabolites in plants is species-specific, and biosynthesis is strictly regulated by developmental stage, tissue or cell group, and by environment (Arbona et al. 2013). While proteins are responsible for the functioning and regulation of all biological processes, proteins serve as the basis of the tight homeostasis that characterises any biological system. Therefore, any changes of secondary metabolites in an organism due to biotic/abiotic stress often correlate with protein synthesis and modification. To elucidate the biosynthetic pathway leading to secondary metabolites, proteomics (the systematic analysis of the proteins expressed by the genome) has been recently used, as many proteins contribute to a cascade of reactions leading to metabolite synthesis and their catalytic enzymes, transport and regulatory proteins (Martínez-Esteso et al. 2015).

Many metabolites present in *C. hindsii* are found in medicinally important chemicals such as drugs, antioxidants, flavours, fragrances, dyes, and insecticides. However, the yields of such commercially important compounds are generally low and cannot meet the increasing demand of the global market (Rai et al. 2017). In addition, there are many tightly regulated steps involved in the biosynthetic pathways of these compounds (Figure 1.1). Therefore, a comprehensive picture of this process has been poorly understood (Jacobs et al. 2000). Hence, to increase plant yields or develop microbial systems to produce bioactive metabolites originating from plants, several alternative strategies have been developed, such as the production of plant cell suspension cultures and metabolic engineering (Facchini et al. 2012). In these cases, it is necessary to have prior information of the proteins involved in metabolite biosynthesis. For instance, Oldham et al. (2010) was successful in identifying proteins involved in the benzophenanthridine alkaloids in *Eschscholzia californica* using a shotgun proteomics approach on yeast-elicited suspension cultures. Moreover, 22 enzymes involved

in monoterpenoid indole alkaloid biosynthesis and 16 proteins predicted to be transporters were identified on cultured *Catharanthus roseus* cells (Champagne et al. 2012).

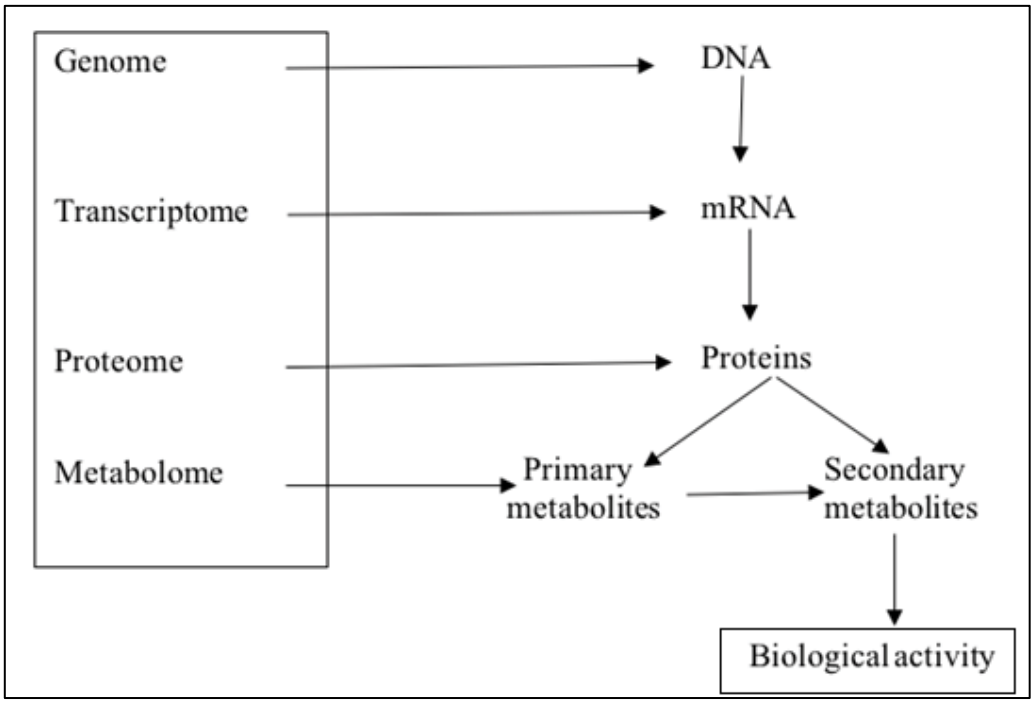


Figure 1.1 Nucleic acids and proteins, and their roles in studying cellular processes

Stress-tolerant phenotypes in crops are the result of differentially expressed proteins that protect them from stresses (Aghaei & Komatsu 2013; Ghosh & Xu 2014; Fu et al. 2016). These protein biomarkers of a given stress response can be potentially modified to enhance a crop's abiotic stress tolerance. In such studies, differentially expressed proteins are identified by comparing stressed crop plants against control ones. Correlations between changes in protein accumulation and the phenotypic response may be elucidated (Abreu et al. 2013; Vanderschuren et al. 2013). Therefore, identification of stress proteins can potentially be used for crop improvement and breeding (Catusse et al. 2011; Gong, Hu & Wang 2015).

To identify proteins involved in metabolite biosynthesis, a differential display analysis has been widely applied to distinguish two or more physiological states of a biological system (Jacobs et al. 2000; Rai et al. 2017). This process consists of four steps: sample preparation, protein separation, identification, and function analysis. To date, 2D-PAGE is still the method of choice as it can separate thousands of proteins from a complex mixture in one experiment. Mass spectrometry (MS) is the most applicable method to identify proteins (Martínez-Esteso et al. 2015). Recently, advances in MS-based protein identification and computational analysis allow the identification of nearly all expressed proteins (Hu et al. 2015). While molecular mass (Mw) and pI of a protein spot may not be enough for protein identification,

the addition of peptide fingerprinting and sequencing can be used in database searches, for example, the bioinformatics tools available from the ExPASy proteomics web server (Gasteiger et al. 2005).

1.2 Summary

The phenotype of an organism results from the combination of multiple interactions among different elements, such as DNA, proteins and metabolites with the environment (Abreu et al. 2013). Therefore, some molecular biology studies integrate information from genetic, metabolite and protein analyses (Nakabayashi & Saito 2015; Rai et al. 2017; Kim et al. 2016; Hall et al. 2011; Afendi et al. 2012). It is notable that a sum of each analysis does not guarantee to produce a clear picture of the actual phenotype, but a sequential characterisation of each element one by one may be necessary. By following this practical framework, we are also able to understand how plants respond to biotic/abiotic stresses (Deshmukh et al. 2014; Nakabayashi & Saito 2015). Relationships among morphological, biochemical, and genetic variation have been studied at different levels in a very few numbers of plants such as *Ocimum gratissimum* L (Vieira et al. 2001), *Lonicera aponica* Thunb (Yuan et al. 2012), *Clitoria ternatea* L. (Ali et al. 2013), and *Withania somnifera* L. (Khanna et al. 2014a)

When a plant has been recognised as an important medicinal plant, especially for commercial purposes, it is essential to identify which phenotypes and geographic variants should be paid more attention to conserve and use for crops improvement and breeding. The critically endangered species *C. fugax* may possess pharmaceutical properties as do other species in this genus (Bamigboye, Tshisikhawe & Taylor 2017; Shiba et al. 2016). However, there has never been a systematic characterisation of morphological, genetic, and metabolite variation and the response of the proteome to such variations in this species. Therefore, it is worth implementing an integrated 'omics' study for this species and *C. hindsii* This will clarify intraspecific genetic variation and assist conservation of the two species, especially to save *C. fugax* from extinction.

The diverse range of antioxidant and secondary metabolites compounds in *C. hindsii* may be valuable for classifying intraspecific variation by analysis of the proteins involved in secondary metabolites biosynthetic pathways, particularly flavonoids and other important antioxidants. Results from this characterisation may help determine how well species, phenotypes and geographic variants validate existing classification. Also, this work can help to reveal information about economically useful genetic variability for researchers, conservationist and plant breeders, and may explain the variable medicinal effect between phenotypes and geographic variants of each species.

1.3 Objective and scope of the study

The principal objectives of the present study focused on studying the genetic diversity of two rare and endangered Vietnamese plants using molecular biology, to use these data to inform conservation decisions, and then to

identify important bioactive compounds present in these plants. Specific objectives include:

1. Determine the most appropriate individual plants or plant populations from which to maintain genetic diversity in both the plant species.

Two methods to measure genetic diversity were used, RAPD-PCR and RAMP-PCR. Data were used to determine heterozygosity in populations. Genetic diversity indices were generated from the data. Two species were examined: *Cycas fugax* and *Celastrus hindsii*, both endangered, rare or threatened plants.

2. Quantifying antioxidants from selected samples of C. hindsii and C. fugax in an attempt to characterise important medicinal properties.

Different protocols were used to quantify the concentration of important antioxidant metabolites and enzymes. Spectrophotometric assays were used to measure the quantified amounts of each compound.

3. Determine up-regulated and down-regulated proteins in high antioxidant (BL) and low antioxidant (NL) C. hindsii, and identify proteins related to pharmaceutical properties.

Protein were extracted and separated by isoelectric focusing (IEF) and subsequently by polyacrylamide gel electrophoresis (2D-PAGE).

Up-regulated and down-regulated proteins were identified based on their intensity on the 2D-PAGE gels. The selected spots were physically excised from the gels and trypsin digested before analysis by mass spectrometry (MS). Further analysis of identified proteins and peptides by MS were conducted to confirm protein identification and function by available bioinformatics tools.

4. Suggest possible technological and management solutions for the conservation of the two species, particularly to improve medicinal plant cultivation as an industrial crop.

Use information available to select the most appropriate plants and populations (selected by genetic diversity and metabolite investigations) and determine a best management for conservation, utilisation, propagation and cultivation (e.g. available resources from local authorities, conservation policy, etc.). Information from the proteomic study may be valuable to develop technical solutions to improve the pharmaceutical properties of plants in both *in vitro* and *in vivo* experiments (e.g. introduction of stress-induced biosynthesis of secondary metabolites, metabolic engineering, etc.).

1.4 Principal questions

(1) Are there particular plants or populations that should be selected to maximise the genetic diversity and sustainability for conservation and sustainable management?

> *This question was answered through an examination of the DNA characteristics and genetic diversity of the two selected species.*

(2) Are the identified anti-oxidative metabolites significantly different between samples/phenotypes/geographic variants of species?

> *This question was answered using metabolite analytical methods to quantify anti-oxidants and enzymes in BL and NL phenotypes.*

(3) What are the proteins associated with the regulation of medicinal metabolites, antioxidants and plant growth and development processes?

> *This question was answered by conducting a differential display analysis to identify up-regulated and down-regulated of proteins from BL and NL samples, and which might have roles in secondary metabolite biosynthesis pathways, growth and development of this medicinal plant.*

(4) What is the pharmaceutical importance of proteins present in each plant extract?

> *Using computational tools and web servers to search, analyse and characterise possible roles of identified proteins and peptides in medicine.*

1.5 Significance of the study

1.5.1 Scientific

The integrated 'omics' study of the two species will:

1. Provide additional information to aid future research and preservation of endangered plants and medicinal plants in general;
2. Contribute and validate an understanding of the relations among morphological, genetic variation and biosynthesis of secondary metabolites of the species.
3. Establish a method for identifying the most appropriate population or individuals from which to introduce genetic diversity for conservation.
4. Validate and contribute to general understanding and applications of the analytical methods used in genetics, metabolomics, and proteomics.

1.5.2 Practical

1. Contribute to the conservation of the two plant species *C. fugax* and *C. hindsii* in Vietnam by identifying priorities for their protection and management.
2. Select the plant phenotypes and geographic variants which possess traits for medicinal purposes, and improve their use as industrial crops, in seedling gardens, botanic gardens, and communities

3. Suggest technological interventions during plant cultivation for medicinal plant crops and highlight scientific research to increase biosynthesis and synthesis of bioactive compounds.
4. Assist in the development of a comprehensive plan for the commercial use of *C. hindsii*, and produce more genetically robust populations of *C. fugax* in Vietnam.

CHAPTER 2

LITERATURE REVIEW

2.1 Plants and conservation

2.1.1 General

Plants are critical for human survival and well-being. Plants provide food for humans and livestock, and a range of other essential products and services. While most commercial plant products come from a limited number of plant species, wild plants contribute to nutrients and food security, as well as traditional medicine systems. Plants are essential for functioning of terrestrial ecosystems (Corle 2016).

According to the updated Global Strategy for Plant Conservation (2010), there are an estimated 500,000 terrestrial plant species, including angiosperms, gymnosperms, ferns, lycophytes, and bryophytes, the majority of which are distributed in the humid tropics, such as the Neotropics and the Asia-Pacific region. Many species of plants are yet to be identified (Pimm & Joppa 2015; Von Konrat et al. 2014). Fewer than 20,000 plant species have been formally assessed by the IUCN Red List criteria (Corle 2016), and the number of plants that are threatened or endangered is not accurately known. Pimm and Joppa (2015) have estimated that a third of all angiosperms are at risk of extinction. The current IUCN Red List (Version 2017-1) of Possibly Extinct in the wild includes only 153 plant species, of which almost one-third survive in cultivation.

Corle (2016) summarised that among five major threats, habitat loss and associated fragmentation causes the most severe depletion of plant diversity, especially in tropical areas where forests are converted to pastures and commercial crops (Steege et al. 2015). Overexploitation of the whole plant or parts reduce the chance of plant survival, particularly for those that have a restricted range or are of high value or rarity. Invasive species threaten native plant diversity, air pollution and nitrogen deposition affect plant biodiversity as plants are exposed to pollutants, particularly in much of Asia (Corle 2016). The final cause of loss in plant diversity climate change. Although there have been no total extinctions recorded, local extinctions, particularly in the climatic margins of species distribution have been documented (Buse et al. 2015).

Conservation efforts for plants, especially threatened and medicinally important species, have been implemented across many nations from the tropics to temperate. Most conservation programs involve improving protected area systems, controlling overexploitation and developing *ex situ* conservation (Corle 2016). These programs require tremendous work in inventory and conservation status assessment before their implementation. Of these, the *ex situ* conservation of tropical plants remains a major gap in global plant conservation because they do not necessarily preserve the full genetic diversity of a species. Therefore, for an effective conservation program, genetic diversity

information needs to be provided, and this can be assisted by molecular techniques (Wee et al. 2015).

Located in the Indo-Burma Biodiversity Hotspot, Viet Nam is ranked as the 16th most biodiverse country in the world. It is home to more than 42,900 identified species, nearly 14,000 of which are recognised species of flora, with a remarkable number of rare and endemic species (Ministry of Natural Resources and Environment 2015). The climate is sub-tropical, tropical and humid, but conditions vary considerably due to the long spread of the country from north to south, with a wide range of topography. The greatest temperature variations are in northern and highland areas. The temperate agro-ecological environment in the northern lowlands is cool in winter months, but in the highlands temperatures are low all year. The tropical climate is observed in the north (in summer), but in the south and the centre, it persists throughout the year. Therefore, there are many wild or cultivated plant species native to different climatic areas.

Out of the diverse flora reported to occur within Vietnam, over 1800 are known to have medicinal uses, and Vietnam ranks the ninth most medicinal plant rich country (Kurian 2012; Uprety et al. 2012) This number may rise to over 4,000 according to a recent report (Hung & Chi 2014). Traditional medicines based on medicinal and aromatic plants have been used widely by all cultural groups and tribes. With its abundant indigenous plant varieties, medicinal plants, herbs and associated traditional knowledge, Viet Nam's biodiversity has a crucial role in contributing to sustainable livelihoods and health care, especially for ethnic minority groups (Dung & Loi 1991; Hung & Chi 2014).

2.1.2 Medicinal plant conservation

Globally, medicinal plants are valuable sources of vitamins and metabolites which play an import role in growth and wellbeing (Aguilar-Støen & Moe 2007). According to plant taxonomists, medicinal plants are present on the entire surface of the earth, of that 33 % are from trees, 32% from herbs, 20% from shrubs, 12% from climbers and 3% are identified as smaller plants (Kurian 2012). They have been used as sources of direct therapeutic agents and a source of new synthetic compounds (Gurnani et al. 2014). According to the International Union for Conservation of Nature (IUCN) and the World Wildlife Fund (WWF), more than 50,000 flowering plant species have been used for healthcare purposes worldwide (Chen et al. 2016). To be a medicinal plant, at least one of its parts (leaves, stem, barks or roots) can be used for therapeutic purposes. Because of availability and relative cost, medicinal plants are often more attractive as therapeutic agents compared to modern synthetic medicines (Banothu et al. 2017).
In Europe, as many as 1300 medicinal plants have been used, and 90% of them are sourced from the wild. In the United States, almost 80% of the 150 prescription medicines are made from natural resources (Chen et al. 2016). In China, traditional medicine accounts for approximately 40% of all healthcare products. It is estimated that 80% of the population in developing countries rely on medicinal plants, and their traditional medicine knowledge for their primary medications, but also for income generation and livelihood improvement (Uprety et al. 2012). Over 25% of prescriptions in developed countries are

extracted from wild herbal plants (Aguilar-Støen & Moe 2007; Abbott 2014; Chen et al. 2016). Recently, while the demand for natural products such as vitamins, antioxidant drugs, and secondary metabolites has been high, the use of the medicinal plant is increasing dramatically all over the world (Li et al. 2015) and Vietnam is not an exception.

Medicinal plants have been used widely for traditional medicines in Vietnam. The practice of using medicinal plants is deeply rooted in culture and is valued, preserved and respected by all groups of people from high income to ethnic minority groups (Hung & Chi 2014). The products of traditional plants for medicinal purposes include active compounds and botanical oils (Table 2.1) for treatment of chronic diseases, functional impairment, and disease related to liver, diabetes and high blood pressure (Woerdenbag et al. 2012; Kurian 2012). Demand for traditional medicines and functional food sourced from medicinal plants has been rising.

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

No	Compound	Plant species	No	Compound	Plant species
1	Rutin	*Sophora japonica*	7	Strychnine	*Strychnos spp.*
2	Berberin	*Coscinium fenestratum*	8	Quinine	*Cinchona spp.*
3	Balmatin	*Fibraurea tinctoria*	9	Scopolamine	*Datura spp.*
4	D-Strophantin	*Strophanthus divaricatus*	10	Beta-Carotene	*Momordica cochinchinesis*
5	Rotundin	*Stephania spp.*	11	Curcuminoid	*Curcuma longa*
6	Mangiferin	*Mangifera indica*	12	Steviosid	*Stevia rebaudiana*

Most medicinal plant species utilised for traditional medicine in Vietnam are collected from the wild in a wide range of habitats, with some large scale trades (Hung & Chi 2014). According to the Country Status report on medicinal and aromatic plants in Vietnam (2014), 206 wild species provide about 12,000 tons per year, whereas there are 53 medicinal plant species cultivated commercially for medicinal purposes. These are mostly grown in home gardens, but some are planted in a more organised fashion, such as sole cropping or intercropping systems and rarely as plantation crops (Hung & Chi 2014). Because the main collection method is from the wild, this adversely affects sustainable utilisation of medicinal plants by accelerating biodiversity loss and the extinction of endangered plant species (Li et al. 2015).

2.1.3 Concerns to medicinal plant conservation

The depletion of plant species has increased, and accelerated the risk of extinction, estimated to now be 100 to 1000 times higher than natural extinction rates (Stuart et al. 1995; Li et al. 2015; Chen et al. 2016). The reasons for these losses include habitat alteration, overexploitation, climate change, fragmentation and degradation of populations, population bottlenecks, invasion of non-natives and genetic drift (Cardinale

et al. 2012). The most common method for collecting medicinal plants is wild harvesting. This is often the best method to meet the quality requirements of the market for medicinal purposes, because accumulation of secondary metabolites is stimulated in natural environments, and may not be expressed under culture or cultivated conditions (Figueiredo & Grelle 2009). This mode of harvesting results in the destruction of natural habitats, especially when this is uncontrolled and in large quantity (Chen et al. 2016). It is estimated that for about 15,000 flowering plant species that are used globally for medicinal purposes, extinction due over collecting and habitat destruction is a possibility, and that one-fifth of their wild resources have already been exhausted (Larsen & Olsen 2007; Chen et al. 2016). Although these threats have been known for decades, the increasing depletion of medicinal plant species and destruction of their habitats has accelerated, especially in developing countries, including Vietnam (Hanh et al. 2014; Minh et al. 2014; Hung & Chi 2014), India (Harish et al. 2014), China (Li et al. 2015; Chen et al. 2016).

2.1.4 Strategy for medicinal plant species conservation

Numerous studies have been conducted on the conservation and sustainable use of medicinal plants (Larsen & Olsen 2007; Uprety et al. 2012; Kurian 2012; Chen et al. 2016). Most of the recommendations raised by these studies suggested that to achieve adequate long-term conservation of medicinal plants, multiple methods must be considered for implementation. All fall under one of two approaches: *in situ* or *ex situ* conservation. The *in situ* approaches involve action to preserve the plant's native habitat. Wild nurseries are considered the most effective method in this approach, to retain the medical productiveness of plants (Sheikh et al. 2002). The *ex-situ* approach complements *in situ* conservation, particularly for species overexploited and exhibiting slow growth and rarity (Hamilton 2004; Volis & Blecher 2010). Currently, *ex situ* conservation by cultivation through propagules and crop variety breeding is the main production mode to ease conflicts between the decrease of natural resources and the development of market demand for medicinal plants (Hamilton 2004; Li et al. 2015; Chen et al. 2016). Either method may contribute to the sustainable conservation of medicinal plant resources and should combine market demand for clean and secure raw materials, ecological stability, and social benefits, instead of controlling unscrupulous medicinal plant removal (Li et al. 2015).

Wild nurseries, also called natural fostering or semi-imitation cultivation is a method suitable for both production and conservation. In other words, it combines economic benefits and diversity preservation, in a practical manner. This method is for species–oriented cultivation and is usually established in a protected area, natural habitat or an area close to the species' natural habitat (Schippmann et al. 2005). Before organising a wild nursery, there are some difficulties to be considered, such as selecting a site in the native habitat or at least ecologically similar to the original habitat, choosing vigorous seedlings and healthy seeds. The requirements for germplasm resources from the wild populations should include the ability to resist stress, high production, and high content

of active compounds. As a result, an investigation of biochemical, molecular and proteomic properties could be very useful to provide key information before a conservation program is initiated (Li et al. 2015).

Cultivation is often used to ensure species survival and can sometimes provide sufficient planting materials in response to increasingly high demand in the market. It also directly saves highly prized medicinal plant species which have been collected to the point of extinction (Schopp-Guth & Fremuth 2001). Although planting materials may not be as medicinally useful as those from wild sources, they still retain high potency, and can be and stored in seed banks for long-term preservation (Figueiredo & Grelle 2009; Joshi & Joshi 2014). Also, cultivation on a large scale, where the opportunity to use modern scientific techniques is possible, can deal with issues of toxic growing components, pesticide contamination, reduced content of secondary metabolites and misidentified plant species (Raina et al. 2011). New techniques applied in cultivation practices can control and provide an optimal input for plant growth and development, such as water level, nutrients, additives fertilisers and even the environmental parameters including temperature, light, and humidity (Liu, Yu & Chen 2011; Wong et al. 2014). Furthermore, with a large amount to supply to the market, planting medicinal materials can reduce production costs and decrease pressure on wild harvesting (Uprety et al. 2012).

However, there are some drawbacks to cultivation when growers cannot afford land, labour, capital and other cropping establishment and management inputs, or deal with environmental variable that affects growth. When cultivation is not properly conducted, benefits to farmers will be reduced, and collection from wild sources may be still carried out. Concerns over the content of chemical substances in cultivated plants still exists (Larsen & Olsen 2007). Even when the cultivation practices are well-conducted, this method might narrow genetic diversity in the gene pool of wild populations and introduced plants are likely to cause genetic pollution of wild resources (Liu, Yu & Chen 2011; Raina, Chand & Sharma 2011; Li et al. 2015). That means a proper selection of the original medicinal plant population is an important prerequisite.
The most difficulty in selecting medicinal varieties is interspecies variation. Different morphological classifications are divided according to the variation including subspecies, variety, and variant, which can be the cause of quality differences between herbal medicines, and different clinical efficacy (Chen et al. 2016). It is difficult to find a good variety possessing not only the highest yield but also the best quality, especially if there is no molecular information. This biochemical and genetic approach must examine multiple characters including habitat specificity, distribution range, population size, species diversity, enzyme and metabolite activities and protein expression of all variants.

2.1.5 Conclusion

It is well-known that geographic and climatic conditions affect the capacity of a medicinal plant species to express its therapeutic properties (Oleszek et al. 2002). Consequently, many researchers have studied geographical variation at the genetic level. The relationship is also significant in designing conservation strategies for plants. Also,

biochemical analyses are valuable methods for taxonomic differentiation within species under different environmental factors (Kramer & Havens 2009). Of these analyses, data from plant metabolomics has been used to discriminate individual species within a genus (Tianniam et al. 2010), and also to identify the original location of particular species (Kang et al. 2008). In other words, the variation in enzymes and metabolites in a biosynthetic pathway may be a consequence of the interaction between genetic and environmental factors, which then contributes to phenotypic variations in the plant. Therefore, when combining conservation genetics with data from other disciplines, supplementary data such as ecological, biochemical, proteomics and cultivation, a multi-disciplinary approach gives a comprehensive picture of plant life and may be sufficient for species conservation, especially for plants having importance economically and medicinally (Khan et al. 2012).

2.2 Genetic diversity

2.2.1 General

The term biodiversity is defined as the variation present in all species of living organisms, their genetic material and the ecosystems in which they occur. In other words, biological diversity occurs at three levels: genetic diversity (variation in genes and genotypes); species diversity (species richness); and ecosystem diversity (communities of species and their environment) (Corle 2016). Genetic diversity is recognised as essential to sustainable development by providing social and economic benefits (Rao 2004). As a major component of biodiversity, genetic diversity represents the heritable variations which can be passed on to offspring and occur within and among populations of a species. Such genetic diversity is encoded in the DNA of an organism and maintained through the major evolutionary processes of mutation, selection, random genetic drift, migration, and mating. In nature, genetic diversity losses often occur as the result of biotic or abiotic stresses, and the survival of wild populations is compromised by reduced genetic diversity (Kramer & Havens 2009). Maintenance of genetic diversity is therefore important for the existence of plant species, as it enables populations to adapt to changing environmental conditions. According to Chevin and Hoffmann (2017), the maintenance of genetic diversity and heterozygosity in natural populations may provide the best general strategy for fitness and long-term survival of most organisms.

a. Factors reducing genetic diversity

Genetic diversity can decline via biotic and abiotic stresses. Extreme natural catastrophes such as floods and droughts may reduce entire populations or reduce the size of populations (Pimm et al. 2014). As a result, genetic diversity stored in the individuals of those populations is severely restricted, and this situation will be more disastrous if the plant species is naturally maintained in restricted habitats. Moreover, further loss of the remaining populations is susceptible to other phenomena such as genetic, demographic or environmental stochasticity (Corle 2016).

Recently, demand on natural products for purposes such as medicine and food has been increasing. Consequently, loss of genetic diversity is accelerating due to habitat destruction, introduction of exotic species, pollution and overexploitation (Kramer & Havens 2009). Of these human-induced activities, overexploitation of rare and endangered medicinal plants whose habitats are small and fragmented is likely to lead to species loss and extinction, as a further consequence of inbreeding depression and accumulation of deleterious alleles (Pimm et al. 2014).

b. Extent and distribution

Genetic diversity can be quantified as the amount of genetic variability among individuals of a variety, or population of a species (Hoban et al. 2013). This variability is a consequence of various genetic differences between individuals and is exhibited in differences in DNA, in biochemical characteristics, in physiological properties or morphological characters. The variation that establishes genetic diversity is derived from mutation and recombination. Understanding the extent and distribution of various aspects of genetic diversity, and its spatial and temporal structure is a crucial prerequisite to determine all aspects of conservation such as what, where and how to conserve it. This situation is urgent and needs to be altered in a way that such genetic information should involve not only describing the variation observed, but also assessing major factors affecting the genetic structure; as well as the effect they may have on the total amount of variation in a plant population (Rao & Hodgkin 2002).

c. Geographic factors

The distribution of genetic diversity is commonly related to geographic differences and isolation. Populations distributed in different geographic locations may differ concerning all aspects of diversity such as the number of alleles, the identity of alleles and their effects on population characteristics (Zhao et al. 2016). The breeding system of the species is crucial in assessing variations among populations from different geographic locations (Rao & Hodgkin 2002). Self-pollinated species exhibit a greater variation between populations, while gradual changes in allele frequency are often found in populations of outbreeding plant species. The pattern is not to assume that geographically close individuals or populations will exhibit genetic similarity or the greatest genetic variation arises from relatively distant populations (Pimm & Joppa 2015). To ensure that the maximum possible genetic diversity is maintained, variation in endangered species should ideally be analysed on an individual basis. Unfortunately, as noted by Narain and Darwin (2000), saving species is expensive, and the funds available for conserving a given species are limited and genetic considerations may, therefore, be placed low on the priority lists of recovery plans (Kramer & Havens 2009). In light of this situation, it would be useful to have a method of analysing genetic diversity that is informative, inexpensive and demands little in the way of chemicals, equipment, and expertise (Corle 2016).

2.2.2 Methods of measuring genetic diversity

a. Genetic diversity assessment

Globally, biological diversity is in decline, and the loss of valuable genetic resources has attracted international conservation action (Corle 2016). To capture maximum variation and minimise genetic diversity losses through time, integrated techniques for germplasm conservation have been developed (Singh et al. 2017). Advances in biotechnology such as *in vitro* culture and molecular techniques are widely applicable for conservation of plant genetic resources (Rao & Hodgkin 2002; De Filippis 2014; Zhao et al. 2016). Assessment of diversity within and between plant populations is routinely performed using various techniques such as morphological, biochemical characterisation in the pre-genomic era, and molecular marker analysis, especially single nucleotide polymorphism (SNPs) in the postgenomic era (Govindaraj, Vetriventhan & Srinivasan 2015; De Filippis 2017, 2018).

(i) Traditionally, genetic diversity variation is characterised by identifying variation in phenotypic traits (vegetative and reproductive organs) or quantitative agronomic traits (yield potential, stress tolerance). This approach, which does not require expensive technology, has certain limitations due to external influences upon morphological characters and quantitative traits (Brown 1978; Andersen & Lübberstedt 2003). This method requires large tracts of land for investigation and experiment and this makes it more expensive than molecular assessment (Mondini et al. 2009).

(ii) To eliminate environmental influences on the characterization of diversity, biochemical methods based on seed storage protein and enzyme electrophoresis have been used and have proven particularly useful in plants (Brown 1979; Ritland et al. 2005) assessed the genetic diversity present within natural forest population of Garry oak. The relationship between domestication practices and genetic diversity reduction of several conifer species have been determined by El-Kassaby (1992); El-Kassaby & Ritland (1996) and Chaisurisri & El-Kassaby (1994). However, their methods are limited by the inability to detect a low level of variation as a result of few isozyme markers with correspondingly low numbers of polymorphic enzymatic systems (Mondini et al. 2009). Thus, the application of isozyme investigation in assessing genetic diversity is quite limited.

(iii) To deal with these problems, the molecular analyses or DNA-based techniques introduced over past decades have the potential to identify polymorphisms with additional advantages over traditional methods. The DNA-based methods have proved convenient by requiring insignificant amounts of plant material which can be collected from any part of the plant. These techniques comprise several types of DNA molecular markers which can be used to analyse genetic variation (Pathak & Abido 2014; Porth & El-Kassaby 2014).

b. DNA marker types and their applications

Comprehensive information about molecular genetic variation present in germplasm plays a crucial role in plant conservation and domestication programs. It would be useful to determine whether morphologically based taxonomic classifications reflect phylogenies. Furthermore, information on the population structure, allelic richness, and diversity of germplasm assists breeders to use genetic resources more efficiently, and help conservationists to select appropriate populations and individuals as the target of conservation (Govindaraj et al. 2015).

DNA markers have been employed widely as an independent method, free from environmental influences (Bracci et al. 2011). They have versatile applications for exploring genetic relationships, diversity, plant breeding, mapping genomes of organisms and building gene bank sequences (Collard & Mackill 2008; Currò et al. 2010). Various DNA markers have been developed for different purposes depending on their features, technical requirements, ability to detect polymorphisms, reproducibility and cost (Semagn et al. 2006).

DNA molecular markers can be classified into three groups based on the method of their detection: hybridisation-based, polymerase chain reaction-based, and DNA sequence-based. The earliest of molecular techniques, namely Restriction Fragment Length Polymorphism (RFLP), was based on restriction-hybridisation techniques with the use of restriction enzymes that recognised and cut specific short sequences of DNA (Mondini et al. 2009). RFLPs have been used extensively to compare genomes in the major cereal families such as rye, wheat, maize, sorghum, barley, and rice (Devos et al. 1993; Bennetzen 2000; Dubcovsky et al. 2001); forest tree (Neale & Kremer 2011). RFLP markers have co-dominant inheritance and have good replication, but are not widely employed due to expense and they are time-consuming (Bardakci 2001).

Later techniques have been based on the use of polymerase chain reaction (PCR), which involves amplification of target DNA sequences. PCR-based methods have become widespread techniques and have led to the development of several types of popular genetic analysis such as Random Amplified Polymorphic DNA (RAPD). For example RAPD has been used to study intraspecific genetic variation in natural population an endangered medicinal plant of *Podophyllum hexandrum* (Nag et al. 2015), genetic similarity among 34 accessions of *Chlorophytum borivilianum* (Tripathi et al. 2012) and genetic diversity of garlic (*Allium sativum*) (Ipek 2003). Based on availability of short oligonucleotide repeats sequences in plants' genomes, different molecular markers have been developed such as simple sequence repeats (SSR), sequence-tagged site (STS), sequence characterised amplification regions (SCAR), and single nucleotide polymorphism (SNP) (Govindaraj et al. 2015; Porth & El-Kassaby 2014). The following sections present the most widely used molecular markers in recent genetic diversity studies on forest and medicinal plants.

c. Random amplified polymorphic DNA (RAPD)

RAPD markers were the first PCR-based molecular markers, developed by Williams et al. (1990), and used for genetic variation analysis generated through the random amplification of genomic DNA using short primers (10 base long synthetic oligonucleotides (Welsh & Mcclelland 1990; Williams et al. 1990). RAPD markers can scan numerous loci in the genome, which make this method suitable for various applications in plant biology such as genetic mapping, taxonomy, population and evolutionary genetics, and plant breeding (Martin et al. 1991; Colombo et al. 2000; Bardakci 2001).

With its favorable properties, RAPD markers have attracted geneticists to construct genetic maps in several plant species such as two *Eucalyptus* species (Grattapaglia et al. 1994), *Pisum sativum* (Laucou et al. 1998), European and Japanese larch (Arcade et al. 2000), white birches (Jiang et al. 2010) two *Dendrobium* species (Lu et al. 2012), and white jute (Chen et al. 2014). To define the species relatives, RAPD markers have advantages of being employed as a method of distinguishing the genetic relationships between families, genus, and species of living organism (Al-Rawashdeh 2011); and were commercially exploited in peanut (*Arachis)* species (Santos et al. 2003).

The most recent studies on the utility of RAPD markers in the phylogeny of the blue-berried honeysuckle gave support to previously taxonomic data of the phylogenetic relationships of *Lonicera venulose* Maxim, *Lonicera boczkarnikowae* and *Loicera caerulea* (Naugžemys et al. 2014). Naz et al. (2014) revealed that 17 commercially important citrus cultivars do not have similar origins as the high value of genetic differentiation among them were examined by using RAPD markers. Using the same technique to research 12 coconut *(Cocos nucifera)* accessions to estimate levels of genetic diversity and structure, Masumbuko et al. (2014) showed different origins for those coconut palms distributed across the coastal belt of Tanzania. By optimizing RAPD-PCR to discover the genetic relationship between *Citrus nobilis* Lour and other species of the *Citrus* genus, as well as their phylogeny, Qian-hua et al. (2011) showed that this species belongs to the tangerine and orange group, and its parents are likely to be either *Citrus haniana* or *Citrus flamea.*

Around the world, thousands of cultivars and wild type populations exist, but limited knowledge is available about their genetic properties. Use of molecular techniques, particularly RAPD markers, by Damasco et al. (1996) on bananas, identified dwarf off-types generated from micro-propagation methods and found that 28% of primers revealed polymorphisms between normal and dwarf plants. The conclusion from the study was that RAPD markers could be used to examine factors impacting tissue culture procedures that lead to the polymorphisms. Furthermore, RAPDs were used to identify changes in genetic diversity following regeneration of potato (Rio et al. 1997) and wheat accessions (Bo̅rner et al. 2000). Studies using RAPD markers were used to evaluate the genetic diversity of different genotypes of plants such as thirty-six pomegranate genotypes among

seven locality groups (Noormohammadi et al. 2012), some species of mahogany (Meliaceae) (Chalmers et al. 1994), and endangered plants (Hung et al. 2011).

RAPD markers have been effective in recognizing genetic variation, and used to find new genotypes in many crops (Desai et al. 2015; Ren et al. 2003; Santos et al. 2003; Zhang et al. 2005; Minh & Tran 2005; Hoang et al. 2009). Genetic variation research on bush mango species used RAPDs to assist domestication programs, and Lowe et al. (2000) recommended that genetic introgression from or into other species probably does not occur after large-scale transplantation. Recently, a study to investigate the genetic relationships among 30 *Brassica juncea* lines and varieties using RAPD markers within "development of canola varieties in *Brassica juncea*" program revealed that breeding of Raya cultivars could be selected efficiently from its recommendations (Tahira & Saleem 2013).

To guide the conservation priorities and strategies for an endangered medicinal plant, *Commiphora wightii,* Harish et al. (2014) used ten each of RAPD and ISSR primers to estimate genetic diversity within and among different populations. RAPD-PCR has been used as the only technique to assess genetic relatedness among 20 mint accessions (Kabir et al. 2014) and investigate the polymorphism among ten elite Indian cultivars of *Cymbopogon* (aromatic grass of essential oil trade types) (Adhikari et al. 2013). However, in the majority of studies, RAPDs were often used in combination with other DNA markers such as ISSR, for example, to characterize genetic diversity of Indian germplasms of *Eauryale ferox* (Kumar et al. 2016), medicinal and commercial plants of *Cymbopogon* (Bishoyi et al. 2016; Baruah et al. 2017), fruit and medicinal trees of litchi (Long et al. 2015), the endangered medicinal plant *Celastrus paniculatus* (Senapati et al. 2013); and combined with AFLP in studies of species of garlic (Ipek 2003). Also, the use RAPD markers have been successful in distinguishing wild and cultivated accessions (Khanna et al. 2014), and measurement of the genetic distance between populations of species (Al-Rawashdeh 2011; Kazemi & Hajizadeh 2012).

The technique of RAPD gained importance because it is simple and efficient (Ren et al. 2003), relatively easy to perform and does not require cloning or sequence information (Huff et al. 1993; Bardakci 2001). Although RAPD has gained popularity because of its advantages, it also has some disadvantages, including poor reproducibility (Neale & Kremer 2011). Earlier limitations of reproducibility in RAPD have been reduced by improved laboratory practices (Cipollini et al. 2013; Ali et al. 2013). Interestingly, the resolution and reproducibility can be greatly increased by an improved RAPD technique (also called RAMP-PCR), in which RAPD primers are used in combination with microsatellite primers (Wu & Tanksley 1993; Fu et al. 2013).

d. Random amplified microsatellite polymorphic (RAMP)

In Eukaryotic genomes, microsatellites or simple sequence repeats (SSR) are widely distributed DNA sequences with short repeated motifs less than six base pairs in length and are informative genetic markers (Min et al. 2008). Sequence tagged microsatellite

sites (STMS) based on SSR DNA sequences have been used as valuable markers for genetic diversity studies in plants thank their hyper-variability, co-dominance, and high reproducibility (Saha et al. 2005). STMS techniques are labour-intensive, time-consuming, and require sequence information before primer design. Thus they have been partly replaced by combinations of SSR with arbitrary primers, such as RAPD primers. RAPD markers have the disadvantage of unable to discriminate heterozygous from homozygous individuals (Bardakci 2001). Therefore, RAPD analysis exhibit more precise estimates between closely related populations and less accurate estimates for distantly related populations in phylogenetic relationships, and may only support existing taxonomy based on morphology (Williams et al. 1993). To compensate for this weakness, a combination of RAPD and microsatellite primers were developed by Wu and Tanksley (1993) and has been proven to be more reliable in relatedness analysis than RAPD or microsatellite methods by themselves (Sánchez de la Hoz et al. 1996).

Like RAPD markers, the RAMP system has proved to have advantages with regards to efficiency, simplicity of performance, applicability with a broad range of primers such as RAPD primers and no requirement for prior sequence information to design of primers (Min et al. 2008). Another advantage of RAMP-PCR is that it is comparable to sequence tagged microsatellite sites (STMS) for detecting polymorphism in genetic diversity studies or molecular mapping (Dávila et al. 1999). RAMP-PCR data are more specific than RAPD-PCR data as the longer SSR based primers, which enable higher stringency amplification, therefore reduces problems with reproducibility, a common criticism of low-stringency RAPD analyses (Yang et al. 1996). By combining SSR and RAPD, RAMP markers have shown better results in cultivar genealogies, being particularly suitable for plant species with an ambiguous genetic background (Zhang et al. 2005).

RAMPs are appropriate for genetic analysis and have been broadly used for genetic diversity and genetic relationship analysis of crops such as wild barley (Dávila et al. 1999), Tunisian fig (Chatti et al. 2007), rice (Hoang et al. 2009a), pomegranate (Zhao et al. 2013a), capsicum (Rai et al. 2013), almond, and apricots (Rasouli & Karimi 2015). Recently, RAMP-PCR has been selected to study genetic diversity of forest and medicinal cultivated plants to support agricultural practices and conservation strategies of *Clinacanthus nutans* (Ismail et al. 2016), *Litchi chinensis* (Long et al. 2015), *Lonicera japonica* (Fu et al. 2013), *Moringa oleifera* (Avila-Treviño et al. 2017). Wei et al. (2016) developed a more effective RAMP-PCR with high–GC content primers (80-100%) to characterise genetic diversity of the medicinal plant *Lonicera japonica*. However, using RAMP the selection of appropriate SSR primer combinations. Therefore the cost to design primer is likely increasing (Min et al. 2008).

2.2.3 Analysis of genetic diversity from molecular data

There are different ways that data can be generated from molecular analysis. Two main types are classified as analysis of genetic relationships among samples and calculation of population genetics parameters. Both these types of analyses start with the construction of a sampling matrix (sample x sample pair-wise genetic distance and similarities).

a. Within-population genetic variation

Genetic diversity analysis can be implemented by measuring the following parameters: the total number of different alleles in the population, the percentage of polymorphic loci, the mean number of alleles per locus, the allelic richness, the within-population genetic diversity (θ), the effective population size (*Ne*), minor allele frequencies, the proportion of heterozygous individuals in the population for a given locus (H_E- the expected heterozygosity), the observed heterozygosity (H_O), and the fixation index (F) (Porth & El-Kassaby 2014).

b. Between/among-population differentiation

Genetic variation between/among (sub-) populations is assessed by the presence of significant allele frequencies differences calculated by various statistical measures such as Nei's gene diversity (H), Shannon's diversity index (I), gene flow (N_m), coefficient of gene differentiation (G_{ST}), genetic distance (D), and correlation between geographic and genetic distance Mantel test (r) (Nei 1973; Jost 2008, 2009; Hedrick 2005). There are two main ways of analysing the resulting distance matrix, namely, principal coordinate analysis (PCA) and dendrograms (or clustering, tree diagram). PCA is used to produce 1, 2 or 3-dimensional scatter plots of the samples. Dendrograms group samples together in clusters that are more genetically similar to each other than to samples in other clusters (Govindaraj et al. 2015).

The molecular data can be scored by the presence/absence matrices manually or with the aid of specific software. However, biases affecting the evaluation process can occur as these scoring techniques are established on the visual detection of bands after gel electrophoresis or else target specific regions in the genome (Rao & Hodgkin 2002b). Table 2.2 presents a brief description of common basic statistical approaches for measuring genetic diversity.

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

Concept term	Description	Formulae/pros/cons
Band-based approach	Scoring presence or absence of banding pattern	Routinely use individuals Totally rely on marker type and polymorphism
Measuring polymorphism	Observing and calculating the total number, percentage of polymorphic bands (PB)	Ib (band informativeness)-ranging from 0-1 according to the formula: Ib = 1 – (2 x (0.5 – *p*) where *p* is the portion of genotypes containing the band
Shannon's index (*I*)	To measure phenotypic diversity and is widely applied	Depend on the extraction of allelic frequencies

Concept term	Description	Formulae/pros/cons
Effective population size (N_e)	Measure the rate of genetic drift, the rate of genetic diversity loss, and an increase of inbreeding within a population	N_e is an idealised number as many calculations depend on the genetic parameters used. Thus, a single population may have many different N_e which are biologically meaningful but distinct from each other.
Heterozygosity (H)	There are two types: heterozygosity observed (H_O) and expected (H_E). Typically values for H_O and H_E range from 0 (no heterozygous) to nearly 1 If H_O and H_E are similar (they do not differ significantly), mating in the population is random. If $H_O < H_E$, the population is inbreeding; if $H_O > H_E$, the population has a mating system avoiding inbreeding.	$H_E = 1 - \sum_i^n p_i^2$ Where p*i* is the frequency of the *i*th allele. H_O is calculated for each locus as the total number of heterozygotes divided by sample size.
G-statistics (Gst)	Genetic differentiation between populations	Measure differentiation well even when heterozygosity is low
Gene flow (N_m)	A standardised measure of genetic differentiation	Related to Gst and has a range from 0 -1 for all loci.
Jost statistic (D)	Jost differentiation statistic	Sensitive to mutation models

2.3 Anti-oxidant enzymes and metabolites

2.3.1 General

Drought, desiccation heat, cold, salinity, extremes of light and various toxic metals are considered the most frequently encountered factors affecting plant growth and development. These stresses influence all levels of the life and development of plants from the morphological and physiological to cellular and molecular (Kidric et al. 2014). Such changes stress conditions require an adjustment of metabolic pathways, aimed at achieving a new state of homeostasis (Shulaev et al. 2008). This re-adjustment process involves the activation of several types of important compounds such as antioxidant, by-products of membrane degradation, and different reactive oxygen species (Shulaev et al. 2008).

Metabolites directly reflect the integration of gene expression, protein interaction and other different regulatory processes in plants (Arbona et al. 2013). Primary metabolites such as sugars and amino acids are direct markers of photosynthetic dysfunction and osmotic readjustment. Secondary metabolites respond particularly to stress conditions, and include antioxidants, reactive oxygen species (ROS) scavengers and regulatory molecules, and are present only incidentally (Edreva et al. 2008; Moore et al. 2014). Regarding their indisputable importance in plant stress physiology, secondary metabolites are alternative defence strategies because plants are immobile and lack an immune system (Edreva et al. 2008). Secondary metabolites are believed to be important for adaptation and survival, responding to the challenges of adverse environmental conditions (Weng 2014).

According to Afendi et al. (2012), there are between 200,000 and 1,000,000 different metabolites produced by the plant kingdom. Many plant secondary metabolites contain bioactive and toxic properties against plant predators, and are also useful for humans to treat various diseases. These specialized bioactive metabolites have been studied to discover disease therapeutics (Pan et al. 2013). Medicinal plants containing a rich varieties of natural products have been described and used in various traditional medicine practices in China, Japan, Vietnam and many other countries (Atanasov et al. 2015). In the last 20 years, more than two-thirds of small molecule drugs have been derived from plant extracts or their derivatives (Facchini et al. 2012).

Some metabolites provide accurate molecular markers for tolerant cultivar selection in conservation and breeding programs (Arbona et al. 2013; Rodziewicz et al. 2014; Nakabayashi & Saito 2015). Profiling plant secondary metabolites gives a better understanding of their stress-protective role, and allows exploration of applications in food, cosmetic and pharmaceutical industries (Shulaev et al. 2008).

Metabolomics includes metabolic fingerprinting, metabolite profiling and targeted analysis. Of these methods, untargeted metabolite profiling is most often used to obtain a global view of the metabolism of cells or identify new metabolites, that can then be validated using targeted quantitative assays (Shulaev et al. 2008; Jorge et al. 2009; Arbona et al. 2013; Deshmukh et al. 2014).

2.3.2 Anti-oxidant reaction

a. Free radicals

Free radicals are atoms, molecules or ions with unpaired electrons. These are highly unstable and active in chemical reactions with other molecules. Oxygen, nitrogen and sulphur are basic sources to create free radicals, which include reactive oxygen species (ROS), reactive nitrogen species (RNS) and reactive sulphur species (RSS). These species exist in biological systems with different states such as ROS including superoxide anion (O_2^{-*}), hydroperoxyl radical (HO_2*), hydroxyl radical (*OH), nitric oxide (NO), hydrogen peroxide (H_2O_2), singlet oxygen ($1\Delta gO_2$), hypochlorous acid (HOCl); RNS formed by reaction of NO with O_2^{-}* creating (ONOO-); and RSS formed by reaction of

ROS with thiols (Lü et al. 2010; Demidchik 2015). The formation of the most popular free radical (ROS) is presented in Figure 2.1.

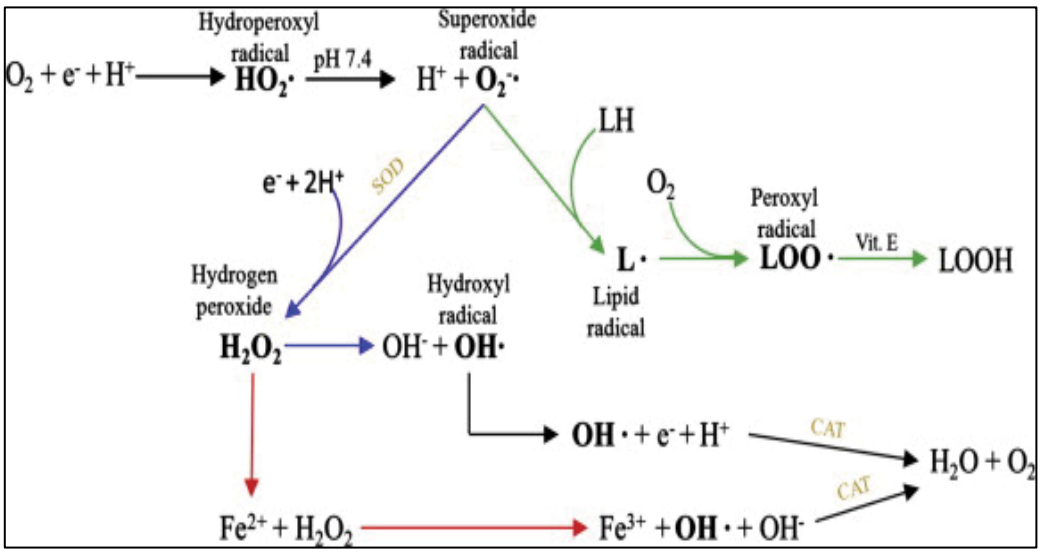


Figure 2.1 Overview of reactions leading to the formation of ROS. Green arrows represent lipid peroxidation. Blue arrows represent the Haber-Weiss reactions, and red arrows represent Fenton reactions. Bold letters: radical and molecules with the same behaviour. SOD: Enzyme superoxide dismutase. CAT: enzyme catalase. Source: Carocho & Ferreira (2013)

Free radicals are produced normally within the plant cytoplasm as a part of metabolism through internal processes like xanthine oxidase, peroxisomes, inflammation, phagocytosis, arachidonate pathways, ischemia (Carocho & Ferreira 2013). These free radicals, although, having an important role in cell signalling, apoptosis, gene expression and ion transportation, will be deleterious if an excessive amount of ROS is created. These adverse effects impact vital molecules including proteins, lipid, RNA and DNA, as ROS chemicals are highly reactive. For example, *OH, a strong oxidant, can attack bases in RNA and DNA, amino acid side chains within protein molecules and double bonds in nucleic acids and fatty acids (Lü et al. 2010, Demidchik 2015).

b. Anti-oxidative metabolites

Antioxidants are able to neutralise free radicals by accepting or donating electron(s) to terminate the unpaired condition of the radical, therefore, stopping the ROS-induced damage (Gülçin 2012). A number of critical studies and reviews have outlined the function of antioxidants and improved understanding of their molecular mechanism of action (Rahman 2007; Krimmel et al. 2010; Tabassum et al. 2010; Lü et al. 2010; Procházková et al. 2011; Terpinc et al. 2011; Noctor, Lelarge-Trouverie & Mhamdi 2015).

Due to the potential advantages in disease prevention and health enhancement, studies on antioxidants have been increasing dramatically over the past decades. The diversity of antioxidant was revised by Carocho & Ferreira (2013) (Figure 2.2). Researchers

have focused on studies of the mechanism of action in biological systems such as cell cultures (Liu & Finley 2005), in clinical trials (Manach et al. 2005), and identified new antioxidants, especially from natural sources (Lü et al. 2010). Antioxidant ability has been measured through total antioxidant activity (Locatelli et al. 2010; Maizura et al. 2010; Saeed et al. 2012) and total antioxidant capacity (Serpen et al. 2012; Wootton-Beard et al. 2011; Floegel et al. 2011; Franzini et al. 2012; Coletta et al. 2014).

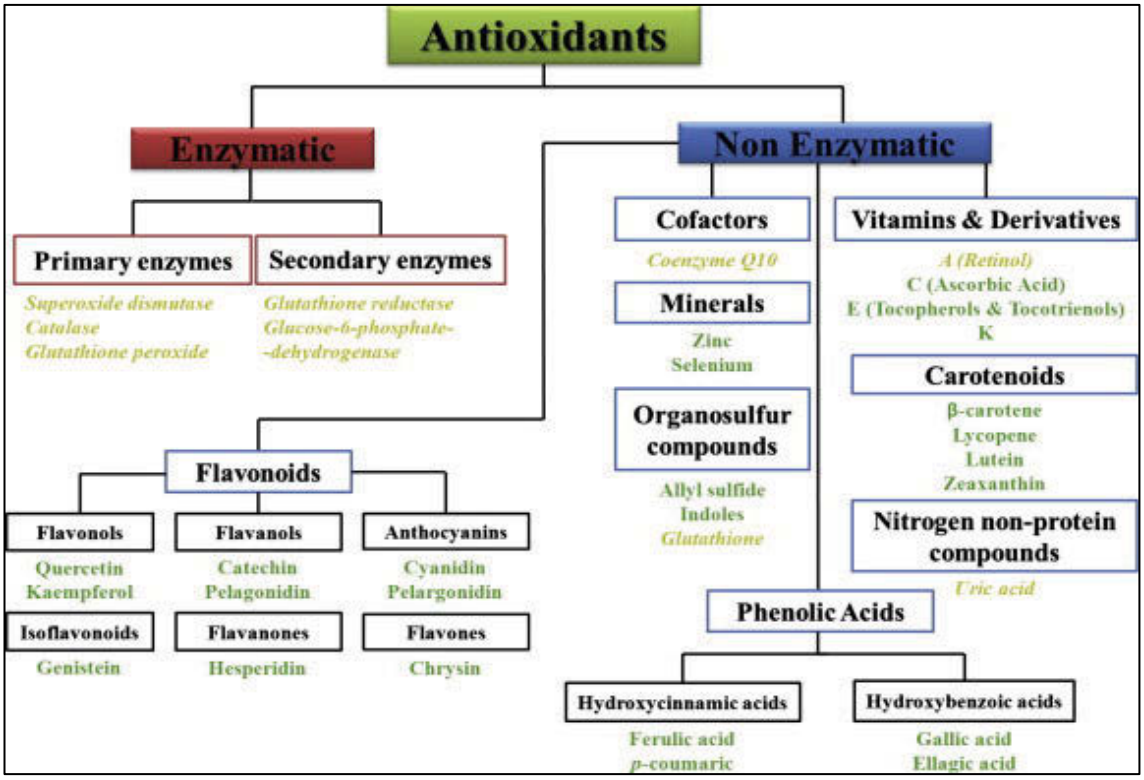


Figure 2.2 Classes of natural antioxidants. Green words represent exogenous antioxidants, while yellow ones represent endogenous antioxidants. Source: Carocho & Ferreira (2013).

c. Antioxidant compounds analysis

To quantify specific compounds, procedures which rely on high-performance liquid chromatography (HPLC), gas chromatography (GC) coupled with other equipment and techniques can be used. Antibody techniques can be employed to identify proteins which are modified by lipid peroxidation products such as the proteins modified within the reaction of unsaturated aldehydes. The fluorescent assay which is used detect aldehydes is based on the principle that aldehydes are modified by polymerisation to form fluorescent products without the presence of amino groups (Gutteridge 1995). Folin-Ciocalteu is a yellow mixture of tungsten and molybdenum, acting as a reagent which reacts with all antioxidants in solution, changing its colour to blue, and therefore, can be analysed easily by a spectrophotometer. However, Folin-Ciocalteu assay's application is limited as it is not selective, while reacting with all reducing species, therefore it is used only as a preliminary approach (Huang et al. 2005).

More recent and modern techniques can examine selected antioxidants. While GC can be used to detect numbers of different compounds, HPLC supported by other detectors has proved to be the most precise separation and quantification method to detect specific antioxidant compounds such as ascorbic acid, flavonoids, phenolic acid, etc. (Carocho & Ferreira 2013).

In short, discovery of bioactive compounds has been advanced through chemical screening, biological screening and high throughput screening. Of these, chemical screening involves the use of latest advances such as thin layer chromatography, high-performance liquid chromatography (HPLC) coupled with detectors and liquid chromatography-mass spectrometry to screen out specific compounds present in the extracts. Whereas, in biological screening or bioassay-guided fractionation, fractions from an active extract are collected in microtiter plates, and each fraction is assayed. These two methods require both time and labour. However, high throughput screening can evaluate many samples in the same biological test sample for their effect on a specific cellular process; as a result, a large number of samples can be tested in a short period and most can be automated.

Chromatographic determination delivers the highest precision and accuracy with superior informative value, but it must be evaluated regarding time and costs of analysis if it is to be used to identify all selected compounds in specific plant material. Furthermore, the measurement of antioxidant activity by HPLC is experimentally more sophisticated (Stratil et al. 2006). On the other hand, it is necessary to analyse samples at least in triplicates as the antioxidant content is influenced by a large number of biotic and abiotic factors. Therefore, using spectrophotometric measurements is, at present, able to obtain sufficient information for evaluation of antioxidants in a comparative biochemical study (Stratil et al. 2006).

In this study, due to limited resources regarding time and labour, high throughput screening strategy will be applied to investigate antioxidant activity of broad and narrow leaf variants of *Celastrus hindsii* using spectrophotometric measurements. Among the species' potential of bioactive properties, only several most relevant enzymes and metabolites will be screened for their antioxidant activities. Protease activity (acid and neutral proteases) will be measure followed by an investigation of several key antioxidants (Glutathione, tocopherols, flavonoids, and phospholipids). Lipid peroxidation will also be investigated through the activity of Malonyl di-aldehyde (MDA). The amino acid proline will be examined to understand the plant response to the environment.

2.3.3 Formation of proteases

Proteases hydrolyse peptide bonds or proteins and break them down into polypeptides or free amino acids (Alnahdi 2012). Within the past decade, there has been much effort to understand the biochemical properties and the mechanism of action of proteolytic enzymes (Palma et al. 2002; Li et al. 2016). This knowledge mostly comes from studies

of human, animal and microbial enzymes, although some studies have been conducted on plants (Sapio & Fricker 2014; Bastos et al. 2017). However, recently more research on plant proteinases has been conducted, providing opportunities to understand their structure and function (Mugford & Osbourn 2010; Wen, Li & Walker 2012; Bienert et al. 2012; Mugford & Milkowski 2012).

The leaves of many plant species contain a rich source of proteolytic enzymes, but the physiological function of the major proteases is not well-known (van der Hoorn 2008). Protease activities in most leaves increases with age and in particular when foliar senescence is promoted by inducing oxidative stress or excision (Simova-Stoilova et al. 2010). Under these stresses, intracellular proteins are oxidised by free radicals and oxidants. These oxidatively modified proteins are selectively recognised and preferentially degraded by intracellular proteolytic enzymes (Simova-Stoilova et al. 2010).

In cereals, grasses and many other herbal and medicinal plant species, such as spinach (*Spinacia oleracea)*, drumstick tree *(Moringa oleifera)*, rice *(Oryza sativa)*, and bean (*Ricinus communis)*, most of the measurable proteolytic activities of leaves are accounted for by a number of acid endopeptidases (Fukayama, Abe & Uchida 2010; Simova-Stoilova et al. 2010; Tajima et al. 2011; Bijina et al. 2011; Maciel et al. 2011). Activities at neutral and alkaline pH have also been detected, in oat leaves for example, where alkaline proteases are located outside the vacuole (Gaur & Wadhwa 2008). Detached leaves of *Lolium* species, which have been extensively used for studies of leaf growth and senescence, accumulate acid protease activity and subsequently replace it with a neutral-alkaline protease in late senescence (Morris et al. 1996).

Protein degradation in plants is a complex process involving a multitude of proteolytic pathways that can be carried out in different cell compartments. The presence of proteolytic activity has been reported in vacuoles, chloroplasts, cell wall, microsomes, mitochondria, cytosol, and the Golgi apparatus. Acid and neutral proteases have complicated mechanisms, and overall knowledge of the mechanisms of action remain to be fully elucidated. Nevertheless, significant advances are being made in understanding the structure, function, and mechanism of proteolytic enzymes (Simova-Stoilova et al. 2010).

In this study, the presence of acid and neutral protease activity in all organelles of the plant under oxidative stress conditions of two cultivars (narrow and broad leaf) of *C. hindsii* will be studied, and the implications of the proteolytic metabolism in different physiological processes will be discussed. Plants have a defence strategy in which various of antioxidant enzymes and metabolites get involved in the reduction of ROS (Kohen & Nyska 2002). It has been confirmed by a range of studies that there is a strong relationship between this defense system and tolerance to biotic and abiotic stresses in a different genotypes of plants (Raza et al. 2007; Bian & Jiang 2009; Maciel et al. 2011; Carvalho et al. 2013; Silva et al. 2016). Many of the enzymatic antioxidants such as superoxide dismutase (SOD), ascorbate peroxidase (APX), monodehydroascorbate reductase,

dehydroascorbate reductase, glutathione reductase (GR), catalase (CAT) and glutathione peroxidase have been studied thoroughly (Choudhury et al. 2016). Other common antioxidants have non-enzymatic activity. In following sections, some non-enzymatic antioxidants including glutathione, tocopherols, flavonoids and phospholipids will be further discussed.

2.3.4 Proline

Proline is a multifunctional amino acid and is essential for primary metabolism. It has been described as acting as an organic osmoprotectant, a metal chelator, an inhibitor of lipid membrane peroxidation (Mehta & Gaur 1999) and as an ROS scavenger (Szabados & Savouré 2010). An increasing number of studies indicate that proline content in higher plants rises under abiotic stresses such as drought (Sharma et al. 2011), high salinity (Kaya et al. 2007; Iqbal et al. 2014), high light and UV irradiation (Aksakal et al. 2017), heavy metals (Singh et al. 2016); and also biotic stresses (Rejeb et al. 2015). Its accumulation normally occurs in the cytoplasm and may be part of a stress signal influencing adaptive responses (Hayat et al. 2012). Also, it functions as a molecular chaperone which can protect protein integrity and enhance the activities of different enzymes (Szabados & Savouré 2010).

However, the correlation between proline accumulation and abiotic stress tolerance in plants varies among species. For instance, proline is accumulated to high levels in *Arabidopsis thaliana* (Liu & Zhu 1997) but found to be not correlated in barley (*Hordeum vulgare*) under salt tolerance (Widodo et al. 2009). In some cases, proline has a toxic effect as over-accumulation occurs, such as the adverse effects found in tomato (Heuer 2003), or poor plant growth as a consequence of high proline concentrations in rice (Roy et al. 1993) and *Arabidopsis* hypocotyl explants (Hare et al. 2001). These studies indicate that proline metabolism in plants has a complex effect on development and stress responses and its accumulation is a result of adverse environmental conditions (Miller et al. 2009).

In plants, proline is synthesised mainly from glutamic acid as a substrate via a pyrroline-5-carboxylate synthetase (P5CS) enzyme and spontaneously converted to pyrroline-5-carboxylate (P5C) (Das & Roychoudhury 2014). Intracellular proline levels are determined by biosynthesis, catabolism and transport between cells and different cellular compartments (Bates et al. 1973), or by activities of enzymes involved in proline synthesis (P5CS, P5C)(Madan et al. 1995) and proline degradation (proline dehydrogenase)(Rena 1975).

2.3.5 Glutathione

Glutathione is the most important low molecular weight antioxidant synthesised in cells. It exists in the thiol-reduced (GSH) and disulfide-oxidised (GSSG) forms (Forman 2016; Winterbourn 2016). This low molecular weight thiol is converted to its oxidised form (GSSG) upon interaction with ROS (Gill et al. 2013). Glutathione serves several vital

functions in physiological activities of the plant, but the first and foremost role is in protecting cells from oxidative damage. It is a main player in removal of many reactive species such as superoxide (O_2) and hydrogen peroxide (H_2O_2) which can lead to the production of toxic oxygen radicals that cause lipid peroxidation and cell injury (Forman, Zhang & Rinna 2009; Lu 2011; Gill et al. 2013; Lu 2014). In this cycle H_2O_2 is reduced to H_2O via ascorbate and reduced glutathione (GSH), and as a result, oxidised glutathione (GSSG) is formed which is recycled back to GSH by the action of glutathione reductase (GR) using NADPH as a reductant (Bhagat et al. 2016). Therefore, it is possible to use GSH as a marker of oxidative stress in plants (Tausz et al. 2004).

One of the crucial issues in measuring the glutathione activity is the measurement of the different forms of thiols in cells. The predominant forms are the reduced forms of GSH and GSSG. Rahman and colleagues (2007) have developed a protocol based on the reaction between GSH with dithionitrobenzoic acid (DTNB), and by reducing GSSG, GSH (GSH + GSSG) can be measured. DTNB reacts with GSH to produce a conjugate and TNB anion that can be detected by fluorescence or absorbance (Rahman et al. 2007).

Other common methods for measuring GSH and GSSG are based on the use of high-performance liquid chromatography (HPLC). For instance, during the measurement, thiol compounds are initially replaced by added iodoacetate. Afterwards, the amino groups on the compound are replaced by 1-fluoro-2, 4-dinitrobenzene. The many compounds produced are separated and identified by their movement on HPLC (Fariss & Reed 1987). GSH is also identified through a direct measurement of nitrosoglutathione which involves in the production of GSH followed by the reaction with orthophthaldehyde (OPT) to produce a fluorescent compound (Tsikas et al. 1999). After reaction with OPT, the products are separated by HPLC with a fluorescence detector (Gladwin et al. 2006). Other rarely used methods include ^{15}N labelling, but this requires the application of mass spectrometry to measure GSH (Kluge et al. 1997).

2.3.6 Tocopherol

While GSH is the most important small molecular weight antioxidant produced in cells, there are other small molecular antioxidants obtained from diet such as tocopherol (vitamin E). Tocopherols have been demonstrated to have a potential health benefit as they can help with hypolipidemic, antihypertensive, allergic dermatitis suppressive, neuroprotective and anti-inflammatory activities (Mocchegiani et al. 2014).

Tocopherols are a major bioactive constituent of human diets and are well-known for their potent antioxidant activities (Saini & Keum 2016). Alpha-tocopherol has the superior antioxidant activity of the tocopherols with three methyl substituents (Seppanen et al. 2010). Unlike GSH, tocopherol is only synthesised by plants and algae and distributed in all parts of plants. In higher plants, α-tocopherols are the predominant isomers contained in the chloroplast double membrane and as a result, are well protected against photo-oxidative attack. Seeds accumulate 10-20 times higher amounts of total tocopherol, with a larger proportion of delta-(γ-) tocopherol (Saini et al. 2014).

Acting as an antioxidant, tocopherol attacks directly the radicals to prevent the chain propagation step during lipid autoxidation. By reaction with lipid radicals in the membrane-water interface, tocopherols donate a hydrogen ion with the formation of consequent tocopheroxyl radicals (TOH^*). The TOH^* radicals undergo recycling to their reduced forms by interaction with GSH and ascorbic acid (Das & Roychoudhury 2014). Furthermore, tocopherol can scavenge oxygen radicals, especially singlet oxygen by deactivating them through charge transfer mechanisms. Apart from its antioxidant functions, tocopherols also play important roles in stabilizing membrane structures, inhibiting cell proliferation or acting as protein kinase C, etc. (Azzi & Stocker 2000; Azzi 2007; Saini & Keum 2016).

Evolutionary, genetic and biochemical evidence shows that tocopherol acts as a factor in cell signalling as result of their important role in maintaining the integrity of long-chain polyunsaturated fatty acids in cell membranes (Azzi 2007). Tocopherol can signal the molecules to change their levels differentially against abiotic stresses (Traber & Atkinson 2007). In short, tocopherols play a governing role in plant stress tolerance (Azzi 2007).

There have been several methods developed to extract tocopherol from biological samples. Tocopherols can be extracted by solvent maceration, with solid-phase dispersion, pressurised liquid, supercritical fluid or ultrasonic-assisted extraction as detailed by Saini & Keum (2016). After extraction, to classify different isomers, the combined tocopherol solution is separated and detected by chromatographic, Fourier transform-infrared spectroscopy and synchronous fluorescence spectroscopy techniques (Saini & Keum 2016). Also, several direct spectrophotometric methods, such as UV-spectrophotometry (Kivcak & Akay 2005) and Raman spectroscopy have been used for tocopherol analysis (Beattie et al. 2017). Each method has advantages and disadvantages, therefore to choose a suitable method it is necessary to take into consideration the physical and chemical status of the sample, and consider the availability of resources and instruments.

2.3.7 Flavonoids

Plants are considered the main source of important pigments such as carotenoids, tetra-pyrrole derivatives, and flavonoids (Stalikas 2007). Of these, flavonoids, the most important secondary plant metabolites, are biosynthetically derived from phenylalanine (Figure 2.3). Flavonoids consist of more than ten thousand substances found in fruits, vegetables and grains, and classified into several subclasses include flavonols, flavan-3-ols, anthocyanins, flavanones, flavones, isoflavones and proanthocyanidins (Merken & Beecher 2000; Welch & Hardcastle 2014; Jiang & Xiong 2016; Cijo, Dellaire & Rupasinghe 2017). Flavonoids are present in nearly every plant; however, each group will be dominant in different types of tissues and organs. For example, fruit with flavonols are commonly found in apples, grapes, and stone fruits, *Citrus* fruits contain flavanones, blueberries, grapes raspberries and cranberries contain anthocyanins, celery, lettuce, and capsicum peppers are rich sources of flavones, and flavan-3-ols are abundant in tea and cocoa (Welch & Hardcastle 2014). Another significant source of flavonoids is from

medicinal plants, and this attracts scientists and the pharmaceutical medicine industry for developing alternative disease treatments (Krishnaiah, Sarbatly & Nithyanandam 2011; Zhang & Tsao 2016; Cijo, Dellaire & Rupasinghe 2017). Flavonoids can benefit human health for allergies, inflammation, viruses, hypertension, arthritis, cancer and HIV (Xiao, Chen & Cao 2014; Alasalvar & Bolling 2015; Zhang & Tsao 2016).

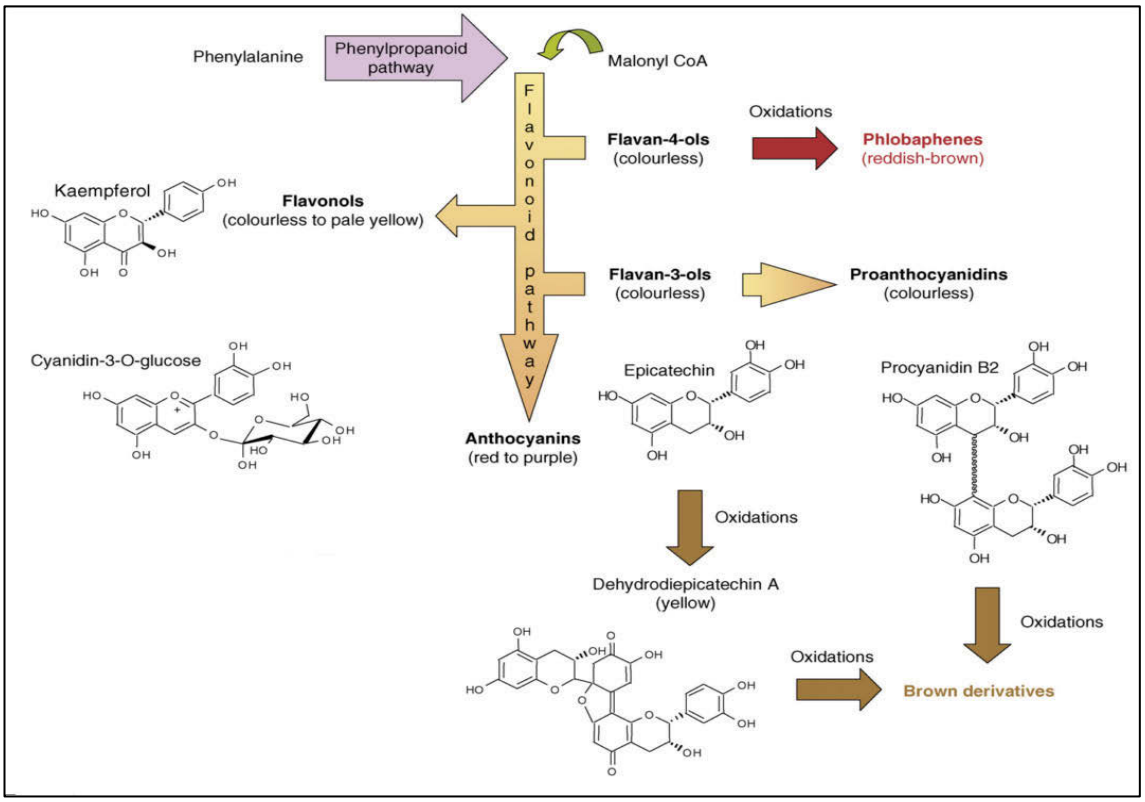


Figure 2.3 Simplified schematic of the flavonoid pathway (*Source*: Pourcel et al. (2007)). This process yields both colourless compounds (e.g. flavanols) and coloured pigments (e.g. anthocyanins).

Antioxidant behaviour is conferred on flavonoids by phenolic hydroxyl groups attached to the ring structures, and these can act as reducing agents, hydrogen donators, singlet oxygen quenchers, superoxide radical scavengers and even as metal chelators (Carocho & Ferreira 2013). Apart from antioxidant properties, flavonoids play other important roles such as activating against other antioxidant enzymes, reducing α-tocopherol radicals (tocopheroxyls), inhibiting oxidases, mitigating nitrosative stress, and increasing levels of uric acid and low molecular weight molecules (Carocho & Ferreira 2013).

Because plants are immobile, they must have a mechanism to protect themselves from adverse environmental factors. Flavonoids act as a secondary antioxidant defence system in plant tissues exposed to different abiotic and biotic stresses. The induced flavonoids are synthesised by plants in response to physical injury, infection or stress (Treutter 2006). Some studies have shown that flavonoids help plants to survive on soils that are contaminated with toxic metals such as aluminium (Barcelo & Poschenrieder 2002), and

have protective functions during excessive light and drought stress (Tattini et al. 2004; Nakabayashi et al. 2014), or frost hardiness (Treutter 2006). Also, flavonoids also regulate growth factors in plants such as auxin levels (Brunetti et al. 2013).

2.3.8 Phospholipids and lipid peroxidation

Lipid peroxidation is a natural metabolic process of fatty acid oxidation that occurs in biological tissues during oxidative stress such as light, heat, or exposure to transition metals or reactive oxygen species (Farmer & Mueller 2013). Polyunsaturated fatty acids, the main components of membrane lipids, are susceptible to peroxidation. The cell membrane will be damaged by free radicals generated from a chain reaction during the complicated process of peroxidation, followed by the disintegration of organelles, oxidation and dysfunction of proteins (Farmer & Mueller 2013). For instance, one of the dangers from lipid peroxidation, besides membrane damage is the production of by-products such as 4-hydroxy-2-nonenal (HNE) which can be broken down yielding a large variety of compounds including unsaturated aldehydes. These aldehydes are toxic compounds because they can react with proteins in the cells, particularly at cysteine, lysine or histidine residues, resulting in inactivation of proteins (Forman et al. 2009).

Plants exposed to various stresses can increase lipid peroxidation as a consequence of the generation of ROS, for instance with *Oryza sativa* under salt stress (Khan & Panda 2008), and *Cicer arietinum* roots under salinity stress (Kukreja et al. 2005). Water stress increases lipid peroxidation and the membrane injury index in leaves of *Phaleolus vulgaris* plants (Zlatev et al. 2006), while Simova-Stoilova et al. (2010) reported that under field drought conditions in wheat plants, membrane integrity was weakened. Sanchez-Rodriguez et al. (2010) investigated the response of five cherry tomato varieties to oxidative stress under water deficit and suggested that lipid peroxidation plays a significant role in selecting tomato plants tolerant to water stress.

The formation of lipophilic radicals within the hydrophobic lipid interior of membranes requires several types of antioxidants for their removal. It is important to retain membrane stability and protection in the presence of phospholipids and cholesterol in a correct ratio (Gutteridge 1995). In lipid peroxidation, phospholipids can act differently such as oxidation substrates or as an antioxidant (Rodriguez-estrada et al. 2011; Cui & Decker 2016).

Phospholipids are main components of cell membranes and are present as a lipid bilayer. They constitute 60% of the lipid mass of eukaryotic cell membranes (Peterson & Cummings 2006) and are sensitive to environmental factors including abiotic and biotic stresses (Cui & Decker 2016). In other words, as a cell signalling function, phospholipids convert external signalling into cells which then transformed their morphological or physiological activities to adapt to changes (Laxalt & Munnik 2002; Xue, Chen & Li 2007). In plants, phospholipids consist of five common forms: phosphatidic acid (PA) which is only found in small quantities in plant tissues; phosphatidylglycerol (PG) mostly distributed in the chloroplasts; phosphatidylcholine (PC), phosphatidylethanolamine (PE)

and phosphatidylinositol (PI) are the main phospholipids of plant mitochondria (Simon 1974; Meijer & Munnik 2003). Recent studies have uncovered that phospholipids are important biomarkers in plant processes such as nutrient uptake, salt stress, heat shock, and stomata movement. (Yu et al. 2010; Lin, Qu & Zhang 2014; Scotti-Campos et al. 2014).

Antioxidant effects on phospholipids have been confirmed in a range of biological samples such as plant oil, animal oil, and marine products (Saito & Ishihara 1997; Cui & Decker 2016). The most general of mechanisms of how phospholipids influence lipid oxidation has been proposed; these are their ability to bind prooxidative metal, produce anti-oxidative compounds, interfere the location of other antioxidants and regenerate primary antioxidants such as tocopherol (Cui et al. 2015). There have been some studies indicated that individual phospholipids possess different degrees of antioxidant behaviour (Saito & Ishihara 1997; Hidalgo, Len & Zamora 2006; Xue, Chen & Li 2007; Cui, Mcclements & Decker 2015). Other possible mechanisms include antioxidant properties of phospholipids to the chelating properties of their primary or free amino head-group component (Dacaranhe & Terao 2001; Cui, Mcclements & Decker 2015), a synergism between tocopherol and phospholipids (Lambelet, Saucy & Loliger 1994; Cui, Mcclements & Decker 2015; Cui & Decker 2016).

Among many techniques used in phospholipid determination, chromatographic techniques appear to be the most useful for isolation and identification of phospholipids mixtures, and NMR spectroscopy is an accurate and precise method for qualitative and quantitative analysis of trace amounts of PL in seeds, oil and lecithin samples. However, regarding the feasibility in time and costs, these techniques are not necessary and spectrophotometric technique can be utilised as a method of choice for measurement of phospholipids in plant tissues (Szydłowska-Czerniak & Szłyk 2003).

2.3.9 Malondialdehyde

Reactive oxygen species (ROS) are highly reactive and attack various classes of biomolecules including protein, acid nucleic and lipids (Farmer & Mueller 2013). For instance, under oxidative stresses, the polyunsaturated fatty acid is peroxidised to finally form malondialdehyde (MDA), 4-hydroxy-2- nonenal (HNE) and other reaction products (Tsikas 2017). While HNE appears to be the most toxic, MDA is the most mutagenic product of lipid peroxidation.

Measurement of MDA in plant tissue is frequently used to indicate the function and amounts of reactive oxygen species. In other words, circulating MDA is one of the most commonly and widely used oxidative stress biomarkers (Valenzuela 1991; Hodges et al. 1999; Farmer & Mueller 2013; Tsikas 2017). Also, MDA has a high capability of reaction with multiple biomolecules such as proteins or DNA, that may lead to the formation of adducts (Blair 2008), therefore identifying MDA production and its role in biology is important to the chemical complexity of oxidative products formed (Ayala et al. 2014). Because MDA reacts with the thiobarbituric acid in an acid medium to give a

characteristic coloured complex absorbing at 532 nm, this has become recognised as a sensitive and specific measure of polyunsaturated lipid peroxidation in biological studies (Gutteridge 1975; Hodges et al. 1999). Recent studies have proposed that data on the formation and concentration of MDA in biological samples may explain both the formation from peroxidation of arachidonic acid and quantitative relationship (Perluigi et al. 2012; Ayala et al. 2014; Tsikas 2017). Other techniques for the determination of free and total MDA, such gas chromatography-mass spectrometry (GC-MS/MS), liquid chromatography-mass spectrometry (LC-MS/MS), and several derivatisation-based strategies, have been developed recently (Ayala et al. 2014). However, the TBARS assay remains popular because of its simplicity, lack of expense, and rapidity in which a large number of samples can be processed with minimal manipulation (Hodges et al. 1999).

2.3.10 Conclusion

C. hindsii are found both naturally in forests and cultivated in home gardens for medicinal purposes. They appear to be morphologically differentiated into a broad leaf isomorph and narrow leaf isomorph variants (see pictures). This morphological difference is perhaps a consequence of different abiotic and biotic stresses and growth characteristics when they are cultivated.

In this study, leaf tissue desiccation and browning were found to be the main obstacle in enzymatic and antioxidant assays for *C. hindsii.* During leaf tissue desiccation, free radicals are key elements causing the metabolic changes including the decrease of chlorophyll, protein hydrolysis, nucleic acid activity alteration, respiratory imbalance, enzymes absorbance and deactivation, polyphenol compounds conversion, changes in fine structure and the breakdown of phospholipids (Noctor et al. 2015). Up to date, there have been only a few studies in the literature of successful investigation of enzymes and secondary metabolites from woody Vietnam medicinal plants, including *C. hindsii* (Ly et al. 2006; Hu et al. 2014). That has been a consequence of both limited research and commercial activity in the area, and the natural difficulties and time-consuming in conducting the assay protocols for woody and high enzymatic activities in plant species. The significance of investigating the degradation processes of proteins, tissue oxidation and membrane lipid peroxidation of plant variants is to better understand the physiology of individual plant species (Demidchik 2015).

2.4 Proteomics

2.4.1 General

a. Introduction of proteomics

To date, there has been no agreement on the birth of proteomics as a recognised discipline, despite the introduction of two-dimensional (2D) gel electrophoresis by Klose (1975). However, many others believe that the term is now more than 20 years old since the term "proteome" was first coined by Wilkins et al. (1996) to circumvent the phrase "all proteins

expressed by a genome, cell or tissue"; and on another level, it may refer to the "PROTEin complement of a genOME". In a broader sense, the definition of the proteome can be presented as being the total set of proteins in a biological entity at all levels from an organelle, cell, tissue, individual and species to the ecosystem at any developmental stage and under specific environmental conditions (Wilkins et al. 1996). By utilising proteomics, the answers for such questions as to how, where, when, and what for are the thousands of protein species produced, the interactions between proteins and other molecules to construct living organisms, that will grow and develop, and adapt with biotic and abiotic conditions. Over the past years, proteomics as a young branch of biological science has evolved slowly at the first stage but rapidly progressed since the human genome sequence was announced, following by the introduction of mass spectrometry, bioinformatics and related equipment. All these developments coupled with improvement in protein isolation methodologies, facilitate the expression of native and engineered proteins (Boughton et al. 2016; Anguraj Vadivel 2015; Jorrin-Novo 2014). The description of the complete proteome of a biological entity, unlike its genome and transcriptome, is not presently achievable, even for the simplest organism. Therefore a comprehensive definition of proteomics is still developing despite the most recent developments (Jorrín-Novo et al. 2015; Finnie 2007).

b. Proteomics development and applications

Recently, life sciences have been attracted enormous attention from scientist to explore biological entities at the systems level (Bantscheff & Schirle 2007). To understand those entities systematically, analytical tools are required to identify the parts of the systems as well as their interactions and responses to changing environment. In response to these emerging requirements, the development of transcriptomic, proteomic and metabolomic profiling techniques have provided substantially better knowledge, especially proteomics which is rapidly accelerating (Bantscheff & Schirle 2007; Bantscheff & Lemeer 2012; Kim et al. 2016). Although many thousand proteomic researches have been conducted and published, a comprehensive picture of all proteins in biological systems is still rudimentary despite the profound influence of mass spectrometry and peptides separation techniques. Bantscheff & Schirle (2007) estimated that proteomic coverage of the genome has rarely achieved 50% and 10% for unicellular organisms and higher organisms, respectively, and significantly less for protein identification. Therefore, a further contribution of proteomics is essential to provide important insights in the biological sciences.

Considered the latest and most advanced stage of biological sciences, proteomics has evolved over the years and contributed to global knowledge of biology. However, the extent of protein information, ranging from a small complex of interacting proteins to subcellular or even whole cell lysates from tissue samples, is largely dependent on the methods used (Ong et al. 2003). In the beginning, there were limited findings on a small number of proteins or gross morphological changes gained from studies of cellular changes. Since the introduction of quantitative proteomics and bioinformatics

advancements, biologists have been able to obtain much more interesting and quantitative data at the molecular level (Bantscheff & Lemeer 2012; Abdallah et al. 2012).

Traditionally, quantitative proteomic methods have been developed by using dyes, fluorophores or radioactivity which has generated acceptable sensitivity, linearity and dynamic range (Bantscheff & Lemeer 2012). These techniques, however, are not able to reveal the nature of the underlying proteins, and often limit their applicability to abundant and soluble proteins, mainly due to requiring high-resolution protein separation. The invention of modern LC-MS/MS techniques and their evolution have overcome most of above problems and opened a new page for proteomics researchers. Mass spectrometry technique, however, creates a wide range of proteolytic peptides, which are physio-chemically differentiated by size, charge, hydrophobicity, etc. (Boughton et al. 2016; Lindemann et al. 2017).

Biologists have been greatly attracted by the broad applicability of proteomics to any part of biology. Most of the early developments were driven by research on humans and yeast. While proteomic studies on humans have focused on searching for biomarkers of disease, notably for cancer, studies on yeast became a model for developing method of choice for other proteomes to deal with the dynamic range of the problems. Although plant proteomics research has been increasing over the past years (Abdallah et al. 2012), the incidence of these studies still lags behind human and animal proteomics, despite the use of some model plant and crops such as *Arabidopsis*, rice, and the availability of their complete genome. Other plant genomic achievements about DNA sequencing projects of corn, tomato, potato, and sorghum are near completion and promise to provide a rich database for plant proteomics (Hu et al. 2015; Jorrín-Novo et al. 2015). It is interesting that most of the quantitative proteomic techniques deployed for human, animal and other eukaryotic organisms can also be applied for plant systems. However, it became more challenging for proteomics as plants have distinct properties including their anatomy, physiology and culture (Champagne & Boutry 2013). Further description of plant proteomics will be presented in the following sections.

2.4.2 Plant proteomics

a. Research on plant proteomics

Plant proteomics was initially not attractive to scientists at the early stage of development until the first plant paper published on rice in *Proteomics* in 2001 by Konishi, Ishiguro & Komatsu (2001) and Imin et al. (2001). To May 2014, there have been 5179 reports published in this journal, but only 365 correspond to plants. According to Jorrín-Novo et al. (2015), the percentage of the total protein solubilised is about 2 - 10% of the total proteome in a single experiment, despite the advancement of modern separation techniques and data searching software. Therefore, some plant proteomic studies will have certain as to what they are identifying without knowledge of the genomic sequence of the organism. Taking all these facts together, it can be suggested that proteins have been the most complex and complicated biomolecules to investigate thank to their instinct properties including chemical complexity, variability, diversity, and dynamic range

(Jorrín-Novo et al. 2015). Despite these challenges, plant proteomics have been studied increasingly and have achieved progress in different fields such as crops improvement (Hu et al. 2015), growth and development processes (Sun et al. 2016; Nawrot et al. 2017), response to environmental conditions and stresses (Ghosh & Xu 2014), organ or subcellular proteome description (Nawrot et al. 2016), PTMs (Canut et al. 2016), translational proteomics, and methodological aspects (Abdallah et al. 2012; Jorrín-Novo et al. 2015). The major aspects of plant proteomics to elucidate include points such as which areas plant proteomics focus on? Which group of plants or their parts? How is biological process getting involved in plant proteomics?

b. Plant proteomics areas

With the purpose of understanding all related aspect of how proteins are formed in living organism, the interaction within proteins molecules and with other molecules, and their adaptation to changing environmental conditions, plant proteomics mainly fall into descriptive proteomics, which characterize all possible proteins species or forms from biological system. Others are comparative proteomics, which have been developed to catalog the differences in protein profiles among genotypes, tissues, organs, developmental stage and external factors (Jorrin-Novo 2014; Jorrín-Novo et al. 2015). However, without additional information from quantitative and translational analyses, proteomics with only descriptive data would not provide confident identification and accurate view and understanding of cell biology. Therefore to minimise inaccuracy, proteomics must have to be viewed as a part of a multidisciplinary approach including other "-omics" research (Rai et al. 2017).

c. Plant species

Plants are common terms for the scientific phyla of *Plantae* when used in searching protein databases. This kingdom is classified into unicellular (green algae) and multicellular eukaryotes which are distinguished as having cell walls and are autotrophic. Most plant proteomics studies have been conducted on multicellular species, including model systems and many others which have agronomic, environmental and industrial importance. According to Jorrín-Novo et al. (2015), 11 species listed account for 85% of the total plant proteomics publications, corresponded to the model systems *Arabidopsis*, *Nicotiana* and agronomic crops such as rice, soybean, pea, tomato, and potato. In contrast, there are few forest trees/woody species and medicinal plants under investigation (Abril et al. 2011; Champagne & Boutry 2013; Aghaei & Komatsu 2013; Jorrín-Novo et al. 2015; Rai et al. 2017). This is probably because the genome of these organisms is not known or poorly sequenced, compared to other experimental model plants and high ranking economic value species. The genomic data can greatly influence the success of proteomic experiments, which require confident protein identification (Rai et al. 2017). In term of taxonomy, most of the plants studied in proteomics are from angiosperms, including either herbaceous or woody species, either mono- or dicotyledonous species.

For non-model species, or so-called anecdotal species, they mostly correspond to orphan species might have great environmental or biomedical value, and with only few protein sequences available in public databases (Champagne & Boutry 2013). Apart from the difficulty in experimental methods, proteomics with these species remains challenging, especially with a database search. While only the use of a Plantae database (NCBI, UniProt) it is not guaranteed to provide confident protein identification, Romero-Rodríguez et al. (2014) suggested that additional information from a custom-built specific database improves the rate and quality of identification in their case studies of *Quercus ilex* and *Pinus radiata*. Others have also recommended the use of expressed sequence tags (EST) databases for protein identification (Konishi et al. 2001; Champagne & Boutry 2013).

d. Plant organs and tissues

During the life cycle of plants, their growth and development are genetically programmed in an environment-dependent manner, creating some different types of cells, tissues, and organs. Proteomics studies recently have included all these levels including the vegetative (stem, leaves or their seedling counterpart hypocotyl, cotyledons, and coleoptiles) and reproductive (flowers, fruits, and seeds) parts. Studies on seeds and leaves, which are fast-growing and photosynthetically active, are dominant amongst publications (Jorrín-Novo et al. 2015). According to Giavalisco et al. (2005), among 2943 protein spots identified in a survey of different organs conducted in *Arabidopsis*, only some of these proteins were organ-specific, while most of them being ubiquitous and corresponding to housekeeping proteins. Therefore, the use of selected plant materials depends on the research questions and objectives of the study. For example, in comparative proteomics, it is likely that only a discrete number of cells respond to the experimental conditions, and if so, it is not necessary to extract proteins from whole organs which might induce high biological variability in proteome analysis. It is noted that confident protein identification will not be obtained from just statistical analysis of the variability. However, not only do plant materials decide biological variability but also greatly affected by growth conditions. This is true in a study on leaf tissue of the field plant *Quercus ilex*, which showed 59% of the biological variability in protein abundance, and 24% due to analytical variability (Romero-Rodríguez et al. 2014). Under controlled growth condition, *Medicago truncatula* showed 25% biological protein differences, and 18% were a result of analytical variability (Asirvatham et al. 2002).

A special case of plant proteomics dealing with homogeneous types of cells such as *in vitro* grown calli, cell suspensions, protoplasts and meristems, which are mostly undifferentiated are not very valuable to represent studies of developmental programs and responses to stresses. However, they are an excellent experimental system for basic research and providing essential information for plant propagation, genetic engineering and bioreactors for producing bioactive compounds (Stone et al. 2017).

It is also necessary to understand the limitations that organs and tissues can impose so that an appropriate protocol can be correctly selected. Roots usually have lower relative

protein content but contain higher salt amounts which make it a hard organ to purify and conduct a proteome investigation. Seeds are popular in proteomics due to their agronomic values, but present more difficulties compared to other vegetative organs as seeds containing a high concentration of polysaccharides, sugars and lipids, while leaves accumulate higher concentrations of phenolics and pigments (Jorrín-Novo et al. 2015). In general, plant proteomics is more recalcitrant than those in humans, animals, and prokaryotes as the presence of molecules, so-called proteomic contaminants, such as proteases, oxidative enzymes, polysaccharides, lipids, phenolics, and other metabolites are in high amounts. The presence of these substances in homogenization reduces the protein solubilization, alters the protein chemical composition and structure, and interferes with downstream applications. Apart from organ-specific variabilities, different stages of development also influence the process of proteomic investigations in the more mature the organs are, the more recalcitrance and protein content are lower, even to stage of senescence with highest proteolysis activities (Jorrín-Novo et al. 2015).

e. Biological processes

In any integrated plant biology research, the ultimate goal is to characterise the processes, mechanisms, genes and genes products that regulate the growth, development and responses to adverse biotic and abiotic conditions. Also, other morphological, physiological, biochemical, and molecular studies integrated into researches programs contribute to the development of more productive plant biology (Cramer et al. 2013; Pedreschi et al. 2013).

Proteomics studies on developmental processes of plants have been mainly conducted on critical stages of flower, seed, and fruit formation, seed germination, seedling growth, and plant establishment (Lehesranta et al. 2006; Pawłowski 2007; Molassiotis et al. 2013; Bourgeois et al. 2009; Fíla et al. 2012). For forest trees, woody crops, and other recalcitrant species, proteomics have been implemented to understand seed dormancy (Pawłowski 2007), germination (Catusse et al. 2011), wood formation (De Filippis & Magel 2012) and clonal propagation (Bian et al. 2009) in an attempt to develop clonal propagation and breeding techniques.

Plant species, along with their growth and development, must have to adapt and tolerate environmental cues by mechanisms that cope with biotic, abiotic stresses, and stress sensing and perception (stress responses). These mechanisms activate intracellula and intercellular signal transduction resulting in gene expression, reprogramming, and protein and metabolic changes (Yordanov et al. 2000; Suzuki et al. 2014). To understand these mechanisms, proteomics is being used more widely and have made a major contribution to unravelling and identifying genes. These include abiotic stresses, temperature, nutritional deficiencies, heavy metals, plant diseases, oxidative stress; which have been widely studied and reviewed (Konishi, Ishiguro & Komatsu 2001; Pawłowski 2007; Sinha & Chattopadhyay 2011; Wang et al. 2013; Aghaei & Komatsu 2013; Ghosh & Xu 2014; Pichereaux et al. 2015; Jorrín-Novo et al. 2015; Fu et al. 2016; Wang et al. 2016). However, it should be emphasised that proteomics itself does not always provide a

comprehensive picture of how plants respond to stress. That is because of several existing drawbacks during proteomics investigations. For example, in nature the stress is progressive, but the experimental system, in case of comparative studies, is not always adequate, and limited to one or two time points, or more often is designed with sparse numbers of biological replicates. Therefore, concerned with high biological variability, and conclusions based on protein abundance should be cautious and conservative (Asirvatham et al. 2002).

f. Methodological generation

At an early stage, plant proteomics was conducted mainly using strategies based on two-dimensional electrophoresis (2-DE) coupled to mass spectrometry (MS). Therefore, only a minority of plant proteomes have been characterised. More recently, proteomics has experienced an explosion of new protocols with improvements in all laboratory steps (protein extraction, depletion, purification, separation, MS analysis) to final computing analysis (algorithms for protein identification and bioinformatics tools for data analysis, databases and repositories) (Anguraj Vadivel 2015). The most utilised platform in plants, particularly woody plant species have been based on electrophoresis (one and two dimensional) coupled to MS (Abril et al. 2011; Rai, Saito & Yamazaki 2017). On the other hand, LC-based separation techniques of peptides (bottom-up) or proteins (top-down)- MudPIT; second generation techniques for quantitative proteomics including both label (DIGE, ICAT, iTRAQ, SILAC) and label-free protocols remain few in woody plant species research (Abril et al. 2011). In the third generation, selected reaction monitoring (SRM) has been recently reported and used in studies with model plants such as yeast (Picotti et al. 2009), *Arabidopsis* (Taylor et al. 2014; Fukao 2015), model legume and *Medicago truncatula* (Van Ness et al. 2016), and a few other woody plant species (Zulak, Lippert, et al. 2009).

Although plant proteomics has experienced through time and technical improvements, there has been an only small fraction of the cell proteome determined, with the exception of a few model plants (*Arabidopsis* and rice). Even for these two plants, the function of quite some proteins is still not known and left behind for future investigations. All the above limitations are due to technical challenges such as sensitivity, resolution and speed of data capture. For un-sequenced species such as forestry trees, woody and medicinal plant species, proteomics, including protein quantification, PTMs, and interact-omics has been come more challenging (Jorrín-Novo et al. 2015).

2.4.3 Medicinal plant proteomics

Medicinal plants are a rich source of highly diverse specialised metabolites with important pharmacological properties which have been used for thousands of years in various traditional medicines. These specialised metabolites are also believed to be the primary strategy of the plant to adapt to changing and adverse biotic and abiotic stresses (Martínez-Esteso et al. 2015). The most valuable secondary metabolites such as alkaloids, anthocyanins, flavonoids, quinones, lignans, steroids, and terpenoids have found

commercial applications as drug, dye, flavour, fragrance, insecticide, and antioxidants (Jacobs et al. 2000). These metabolites are characterised by complex chemical structures, regulated by biosynthetic mechanisms involved in natural selection (Moore et al. 2014). The development of "omics", particularly proteomic technologies have contributed to the progress made in the characterization of plant secondary metabolism pathways (Table 2.3).

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

No.	Medicinal plants	Identified protein	Biosynthesis pathway of	Reference
1	*Catharanthus roseus*	Strictosidine synthase Tryptophan synthase 12-oxophytodienoate reductase	Strictosidine in alkaloids Tryptamine as alkaloid precursor The regulator jasmonic acid	Jacobs et al. (2005)
2	*Panax ginseng*	Enolase glyceraldehyde 3-phosphate, dehydrogenase, aldolase	Ginsenoside 14 differential amino acid	Nam et al. (2005) Sun et al. (2015)
3	*Chelidonium majus*	Disease/defense-related proteins Nucleic acid binding proteins	Secondary metabolites	Nawrot et al. (2007;2016)
4	*Papaver somniferum*	Condeinone reductase	Morphine	(Decker et al. 2000)
5	*Papaver somniferum*	Sanguinarine	Alkaloid	(Desgagné-Penix et al. 2010)
6	*Eschscholzia californica*	Methionine synthase Phosphofructokinase	Benzophenanthridine alkaloids (S)- adenosyl methionine	Oldham et al. (2010)
7	*Catharanthus roseus*	22 Enzymes	Terpenoid indole alkaloid	Champagne et al. (2012)
8	*Papaver somniferum*	6-O-methyltransferase (6OMT) and 7OMT	Isoquinoline	Ounaroon et al. (2003)
9	*Pseudostellaria heterophylla is*	71 differentiated protein	Amino acids	Hua et al. (2016)

Although proteomics recently has been proven to play an important role in investigating the plant stress responses under different conditions (Aghaei & Komatsu 2013; Gong et al. 2015; Ahmad et al. 2016), its role in our understanding the biosynthesis of secondary metabolites and their regulation in non-model medicinal plant species is quite inadequate; except for major families of metabolites, i.e., phenolics, alkaloids, and terpenes (Rai et al. 2017). Nevertheless, there have been considerable achievements in an attempt to identify the proteins specifically localised within tissues that biosynthesise active specialised metabolites in medicinal plants (Martínez-Esteso et al. 2015)

Proteome analysis using (LC-MS/MS) coupled with transcriptome analysis have revealed the component of alkaloid metabolism in opium poppy (*Papaver somniferum*) cell cultures with the presence of sanguinarine biosynthetic enzymes (Desgagné-Penix et al. 2010). Using

shotgun proteomics approach, Oldham et al. (2010) reported the role of candidate proteins involved in the benzophenanthridine alkaloids on yeast-elicited *Eschscholzia californica* suspension cultures. In the proteomic study on cultured *Catharanthus roseus* cells, 63 enzymes were identified to be potentially involved in secondary metabolism, and of those 22 enzymes were involved in monoterpenoid indole alkaloid biosynthesis. Friso et al. (2010) used high accuracy mass spectrometry and protein quantification by spectra counting to compare the differential bundle sheath and mesophyll cells of maize *(Zea mays)* trying to understand the quantitative distribution of the primary and secondary metabolic pathways of these cells. Recently, Miettinen et al. (2014) implemented proteomic experiments on epidermal, and mesophyll protoplasts isolated from *Catharanthus roseus* leaves to examine the enzymatic biosynthesis and localisation which were then used to assist transcriptome analysis to identify the genes of seco-iridoid biosynthesis. In order to identify proteins involved in flavonoid and phenylpropanoid biosynthesis pathway, Tan et al. (2012) utilised MALDI-MS proteomic technique and reported that 11 identified proteins were related to the biosynthesis of cyclohexenyl chalcone derivatives in cell suspension cultures of *Boesenbergia rotunda*. In other studies, proteomics itself or in combination with other "omics" have identified a number of proteins involved in different biosynthesis pathways. For example, 58 proteins from cell cultures are keys enzymes for the biosynthesis of monoterpenoid indole alkaloids (Jacobs et al. 2005); and proteins involved in biosynthesis pathway of artemisinin (Bryant et al. 2016)

Comparative proteomics analysis also has been applied to identify the amino acid metabolism and metabolite biosynthetic pathway of some medicinal plants when comparing between two or more cultivars (differentiated by morphological characteristics or growing habitats). Sun et al. (2016) used two-dimensional polyacrylamide gel electrophoresis and isobaric tags to identify the differential abundance of proteins between wild and cultivated ginseng and revealed that 14 types of amino acid were higher in wild cultivar which also accumulated the higher level of methionine synthase, glycolysis and tricarboxylic acid cycle related enzymes. In another study, Hua et al. (2016) also identified 71 significant differential expression of proteins in cultivated *Pseudostellaria heterophylla* and its wild-type using iTRAQ-base quantitative proteomics analysis.

Proteomics have brought promising insights into plant secondary metabolism, however unlike other sub-areas under the plant proteomics category, several theoretical and practical issues need to be taken into further consideration in the analysis of secondary plant metabolites such as the proteomic strategy to access target proteins; on which approach the proteomic investigation should follow; and selection of suitable plant material in secondary metabolites of interest. The workflow of a standard MS-based proteomics experiments includes all or most of the following steps: experimental design; sampling tissue/cell or organelle preparation; protein extraction; MS analysis; protein identification; statistical analysis of data; validation of identification (Figure 2.4). All these will be detailed in the following sections.

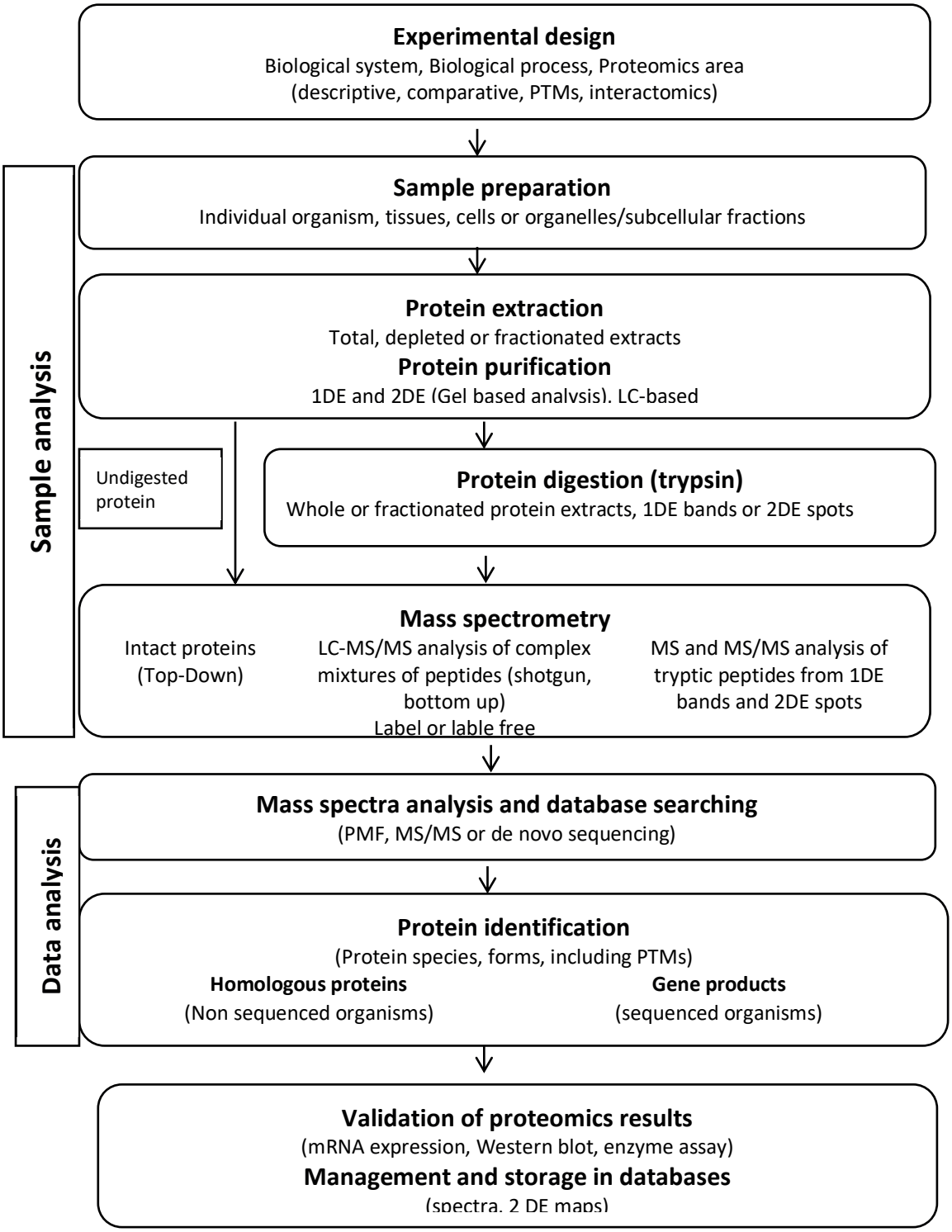


Figure 2.4 Workflow of a proteomic experiment

2.4.4 Proteomic strategy and approaches in secondary metabolism

a. Proteomic strategy

The proteins of interest to secondary metabolites can be identified by a comprehensive or a differential analysis, which are being used in different approaches. While the comprehensive analysis, many proteins intended are prepared for identification, the differential analysis only focuses on differentiated protein abundance between samples. Although comprehensive analysis has been widely applied to identify and characterise protein presented in chromoplasts (Zeng et al. 2011), trichomes (Van Cutsem et al. 2011) or in elicited cell cultures (Zulak, Khan, et al. 2009), it is being used less than differential proteomics as a strategy. That is because of the similar role of mRNA sequencing, which can provide better dynamic range coverage at an economical cost (Zubarev 2013). Up to date the quantification of differences between two or more physiological states of a biological system is one of the most popular tasks but is technically challenging to proteomic biologists (Bantscheff & Schirle 2007).

b. Proteomics approach

Differential proteomics approaches can be classified as either label-base or label-free quantification methods, as well as gel-free or gel-based, of which each approach can be further subdivided into the various types of methods such as chemical and metabolic labelling in label-based approach, and 2-DE, DIGE and iTRAQ in label-free approach. Label-free relative quantification with classic 2-DE continues to be the most common method of choice in the field of quantitative proteomics by mass spectrometry which are designed to find the up- or down-regulated proteins involved directly in the biosynthesis of metabolites (Anguraj Vadivel 2015; Martínez-Esteso et al. 2015). Without the utilisation of labelling reagents, this method makes the quantification simple and cost-effective. The key principle of the label-free method is based on counting the number of peptide-to-spectrum matches obtained for each protein or measuring peptide signal intensities by using the extracted ion chromatogram. In differential display analysis, the integrated signal response of individual peptides is compared between LC-MS (/MS) runs of different samples. Differential protein abundance, as a result, is conducted by an aggregation of variations identified for all peptides matching the respective protein. This method, therefore, can measure and compare an unlimited number of samples and the analysis of untreated proteins or peptides. A few key enzymes directly related to terpenoid indole alkaloid biosynthesis have been detected in the study of two independent, cultured cell lines of *Catharanthus roseus* (Champagne et al. 2012). However, there were no clear results obtained by 2-DE differential analyses while studying secondary metabolite enzymes of Korean ginseng (*Panax ginseng*), Indian ginseng (*Withania somnifera*) (Nagappan et al. 2012) and (*Euphorbia kansui*) (Zhao et al. 2014). The unsuccessful finding may be partly due to the drawback of 2-DE in the detection of a lower abundance protein. To minimise this limitation, Zhang et al. (2014) used the combination of hexapeptide ligand libraries (CPLL) as an enrichment strategy to study the effect of UV

stress to the induction of the terpenoid indole alkaloids (TIAs) production on *Mahonia bealei* leaves.

Other methods that had better perform the differentiation have also been employed in studies of secondary plant metabolites. For example, 2D - difference gel electrophoresis (DIGE) has been used to identify three terpenoid indole alkaloid (TIA) biosynthetic enzymes (Champagne et al. 2012), six flavonoid pathway enzymes presenting ripening grape berries (Martínez-Esteso et al. 2011). In another similar study but using iTRAQ method, Martínez-Esteso et al. (2013) found 38 proteins regulating the shikimate, phenylpropanoid and flavonoid pathways. From Martínez' studies, it may be concluded that iTRAQ is superior to DIGE in proteome coverage. Thus, Hua et al. (2016) used iTRAQ to record the changes between cultivated and wild *Pseudostellaria heterophylla* and detected 71 significant differentially expressed proteins involved in carbohydrate and cellular amino acid metabolism. With gel-free methods, Oldham et al. (2010) used shotgun proteomics combined with a well-annotated, translated unigene database, and reported nine differentially abundant proteins among 646 proteins identified involved in (S)-adennosyl methionine biosynthesis and benzophenanthridine alkaloids production.

While gel-based methods profoundly improve the detection of isoform and PTMs, gel-free benefits the understanding of broader proteome coverage in combination with extensive fractionation (Martínez-Esteso et al. 2015). However, gel-based quantification, in contrast to other labelling techniques, has several technical challenges as each sample has to be run separately, which negatively influences its reproducibility and requires consistent workflows (Lindemann et al. 2017). To minimise the technical variances, caution during all sample preparation steps and LC-MS measurement need to be devised seriously before any projects start. Despite its limitation, it is recommended that 2-DE gels should not be replaced by other methods which may be regarded as superior and complementary. The selection of which approach to be adopted is often determined by the biological question addressed, time, cost, and availability of instruments (Abdallah et al. 2012; Lindemann et al. 2017).

2.4.5 Sample preparation, protein identification and bioinformatics

a. Plant material

The first issue which needs to be decided is to select appropriate plant material (organ or tissues) that contains a rich source of secondary metabolites of interest. It is also important for a proteomic study on potential pathway enzymes to select a suitable plant part (organ or tissues), and either an *in vitro* or *in vivo* experiment, where the target pathway is over-represented. For instance, studies on the pathway of flavonoid and volatile aldehyde synthesis by Schilmiller et al. (2010) revealed that terpenoid metabolites produced by a sesquiterpene synthase were only found at the protein level in leaves of tomato, but not stem or trichomes. The roots of *Panax ginseng* were used in a proteomic analysis of amino acid metabolism (Hang Sun et al. 2016). Proteomics analysis of the medicinal plant *Artemisia annua* was conducted on leaves and trichome extracts and showed that

additional enzymatically driven processes occurred within the trichomes in the biosynthesis of artemisinin (Bryant et al. 2016). However, it is noted that the successful identification of proteins is mainly depended on the availability of sequences in the databases of the selected species. For example, while only four of the 19 identified proteins had been functionally described in species of *Euphorbia kansui* (Zhao et al. 2014), several hundreds of proteins were identified and had functional descriptions in species of opium poppy (Onoyovwe et al. 2013).

On the other hand, *in vitro* cell cultures have been selected in a number of studies as their equivalence with specialised tissues where the metabolic pathways occur under controlled conditions with the induction of stress, precursor feeding and elicitation (Stone et al. 2017). Most of the proteomic studies on secondary metabolites have been conducted with elicited cell cultures of several medicinal plants such as flavonolignan in *Silybum marianum* (Corchete & Bru 2013), lignans in *Podophyllum hexandrum* (Bhattacharyya et al. 2012), and chalcone derivatives in *Boesenbergia rotunda* (Tan et al. 2012). Elicitation and phenotype comparison strategies have been embraced for differential analyses to discover the alkaloid biosynthetic pathway in cell cultures of *California poppy* (Oldham et al. 2010), opium poppy (Desgagné-Penix et al. 2010), and Madagascar periwinkle (Champagne et al. 2012). Despite cell cultures have been the most popular plant material, the metabolic profiles might differ from plant tissues and organs sources as the pathways are large and complex. Thus, this approach is useful for investigating particular parts of the target biosynthesis process including secoiridoid as part of terpenoid indole alkaloid biosynthesis (Champagne et al. 2012).

b. Protein extraction

Major improvements in proteome techniques in recent years have led to an increase in their application in all biological fields, including the plant sciences (Isaacson et al. 2006). For all proteome procedures, protein separation and sample preparation are of importance for optimal proteomic results (Saravanan & Rose 2004; Westermeier 2014; Wu et al. 2014). As plant tissues contain relatively low quantity of proteins and high concentration of proteases and other compounds that potentially affect the quality of the protein extracts and adversely impact to downstream process of protein separation and identification, the ideal procedure should capture all proteins species in a proteome, minimize protein degradation and eliminate maximum contaminants (Alam et al. 2013). A wide range of techniques for protein separation and sample preparation are available based on their physicochemical and structural characteristics; such as solubility, hydrophobicity, molecular weight, isoelectric point (pI), etc. Generally, to obtain the protein fraction of interest, the technique usually consists of the processes of cell disruption, solubilisation/precipitation, and enrichment systems (Martínez-maqueda et al. 2013). A general procedure can be described as in Figure 2.5.

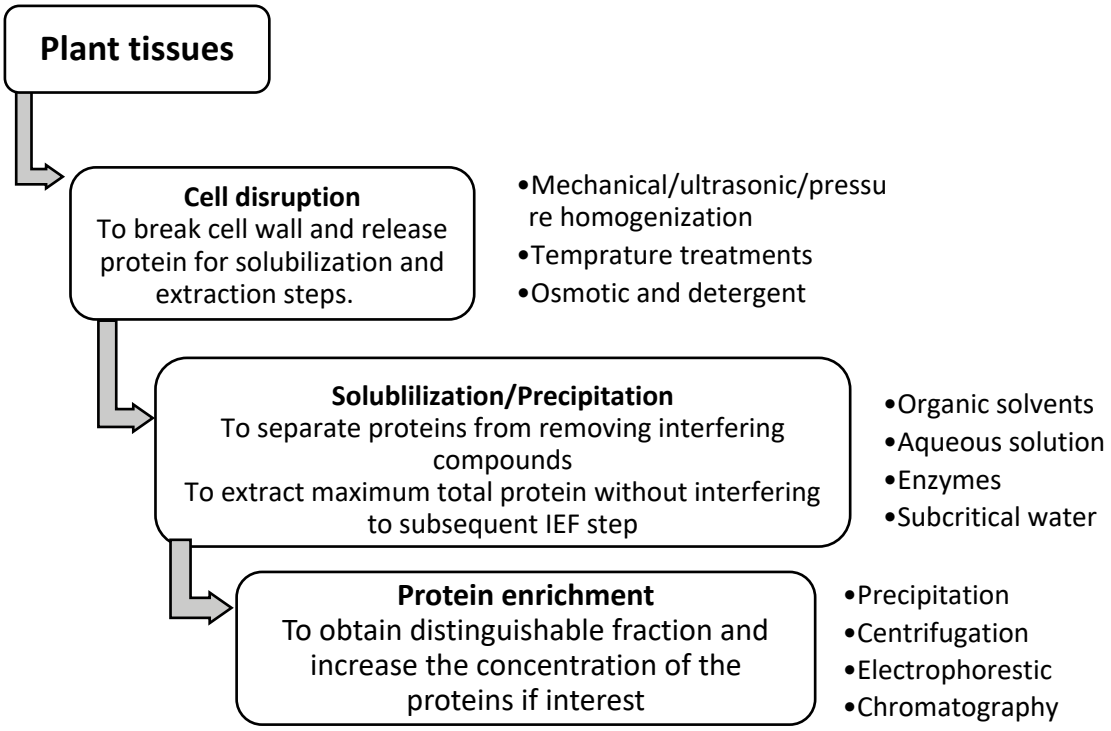


Figure 2.5 The preparation of protein extracts from plant tissues

The protein extraction involves breakage of the cell and release of the contents into a soluble buffered medium. By the breakage of the cell involves physical disruption by mechanical means, sonication, chemical or enzymatic lysis, the target analyte proteins are usually released into a buffered mixture. The addition of other compounds that facilitate solubilizing the protein, can protect proteins from hydrolysis or oxidation, and eliminate non-proteaceous components in the aqueous extracts (Newton et al. 2004). Generally, different non-proteaceous constituents can be removed by several means; lipids removed by detergents, nucleic acids by selective pH precipitation, carbohydrates by specific enzymes, while proteins are protected by antioxidants (dithiothreitol, mercaptoethanol and protease inhibitors (Newton et al. 2004).

(i) Tissue disruption

The disruption of plant tissues become problematic because of the presence of plant cell walls which consist of cellulose and its derivatives, or lignin and related compounds, especially in woody plants (Wu et al. 2014). Also, proteins are usually contained in protein bodies inside cell walls, so cell disruption is an important prerequisite before the solubilisation and extraction of total proteins. Because this step is very important, various chemical and techniques are usually in use for a complete cell disruption such as mechanical homogenization (Boye et al. 2010), ultrasound homogenization (Chittapalo & Noomhorm 2009), pressure homogenization (Dong et al. 2011), temperature treatments (Sheoran et al. 2009), and osmotic and chemical lysis (Doolan & Wilkinson 2009; Kim et al. 2011).

A common plant tissue disruption procedure includes the process of pulverising materials by using mortar and pestle with liquid N_2. In some cases, quartz sand and stainless steel ball are also used to obtain a fine powder from fibre-rich tissues of woody tissue, stem and aged leaf (Kumar et al. 2017). The process of plant disruption is considered tougher than in other organism cells because of the outermost walls, which contains polysaccharides and other compounds. Also, using a small amount of starting materials (e.g. 0.2 – 0.5 g of fresh weight) and the fineness of the tissue powder affect the yield of total protein considerably. As a result, a device generating strong shearing forces such as a cryomill with stainless steel balls is also used to produce a fine protein powder (Kumar et al. 2017).

(ii) Removal of interfering compounds

Plant tissues contain substantial amount of secondary metabolites, including phenolics, flavonoids, tannins, lignins, etc., and nucleic acids. These contaminants severely affect the process of protein extraction and separation through building irreversible complexes with proteins and can cause streaking and art factual spots on 2-DE gels due to the oxidation of phenolics (Wang et al. 2003; Vâlcu & Schlink 2006). Also, while the existence of secondary metabolites varies with age or development stage, especially abundant in adult, green tissue compared with young, etiolated tissues, nucleic acids are more abundant in rapidly dividing cells and young tissues (Wu et al. 2014). Therefore, the selection of tissue for sample preparation becomes critical.

For eliminating these contaminations, a removal process can be conducted before protein extraction (Toth & Pavia 2001; Wang et al. 2006) using 10% w/v TCA/acetone (plus 0.07% 2- 2-ME or after protein extraction with the extraction buffer of EDTA, DTT or 2-ME, a protease inhibitor cocktail, and PVPP. After the "clean-up" process the first strategy, in that the tissue pellet should be white or very lightly coloured to signal that the majority of secondary metabolites have been removed. In the second strategy, the proteins are directly extracted from homogenised tissues in an aqueous buffer. The removal of interfering compounds before protein extraction can substantially facilitate downstream manipulation, the removal of those after protein extraction is more often applied for easy tissues such as young parts of the plant because the co-precipitation between contaminants and proteins are usually difficult to remove in subsequent organic solvent clean-up steps (Wu et al. 2014).

(iii) Methods for protein extraction/purification

Up to date, although many protein extraction methods for plant tissues have been developed for downstream procedures of proteomic analysis, they are generally based on the use of TCA/acetone for precipitation (Granier & de Vienne 1986), phenol extraction (Hurkman et al. 1989; Isaacson et al. 2006). Recently, the method developed by Wang et al. (2007) using ammonium sulphate saturated-methanol and Borax/PVPP/Phenol has been widely used.

(1) TCA/acetone extraction

Since initially developed by Damerval et al. (1986), TCA in acetone has been commonly used for plant protein extraction (Sheoran et al. 2009; Zhen & Shi 2011; Song et al. 2012). The combination facilitates proteins concentration and removal of contaminants due to the principle that protein is denatured under acidic and hydrophobic environments. Using a TCA/acetone technique has several advantages: valuable mean to inhibit the activation of protease (especially at -20^{0}C) and even for subsequent re-solubilisation of proteins; removal of interfering compounds; and enrichment of alkaline proteins such as ribosomal proteins (Saravanan & Rose 2004; Isaacson et al. 2006). This method has proven good on young plants, however, it exhibited limitations in complex plant tissues. The proteins recovered by this method are difficult to redissolve (Yao et al. 2006), and salt ions are hardly removed efficiently (Görg et al. 2004). Nevertheless, it is still recommended a starting protocol for plant protein extraction.

(2) Phenol-based extraction

Since first introduced by Hurkman & Tanaka (1986) to solubilise the membrane proteins of barley root, this method has been widely used for total protein extraction from plant tissues as an alternative to TCA/acetone extraction (Wang et al. 2008). In this method, plant tissues powder is extracted in the extraction solution/buffered phenol (pH 8.0) and followed by methanol (or acetone) precipitation of the phenol phase. Phenol dissolves proteins and lipid, and leaves polysaccharides, nucleic acids (water-soluble substances) in the aqueous phase, then the proteins are precipitated by the methanol (or acetone) (Wang et al. 2003). To deal with plants containing high levels of phenolic compounds, PVPP (0.05g/g tissue) (Wang et al. 2006) is added to the pulverizing step, or/and PVPP (1% w/v) is added to the lysis buffer prior to phenol extraction; which can facilitate the protein extraction outcome (Wang et al. 2007), or adding 2% w/v concentration of sodium dodecyl sulphate (SDS) to the extraction buffer to improve protein solubility (Wang et al. 2003). The major advantages of the phenol-based protocol are: containing toxic chemicals and consumes more time than TCA/acetone extraction method, the pellet is hard to solubilise and this is necessary to consider the sample preparation steps carefully (Chen & Harmon 2006; Wang, Tai & Chen 2008).

(3) Borax/PVPP/Phe extraction (BPP method)

With the integration of phenol-based protocols and TCA/acetone protocols (Wang et al. 2007), the methods exhibit substantial superiority than each method by itself. The combined extraction protocol has been proven to be very effective to deal with recalcitrant plant tissues and facilitate 2-DE-based proteomic analysis by removing most of the interfering compounds and produce high-quality protein samples. Although the Borax/PVPP/Phe extraction cannot eliminate the drawback of toxicity, it is widely used for high-quality protein products extraction (Wang et al. 2013; Zhang et al. 2014; Li et al. 2015). In short, a single universal extraction protocol could not provide sufficient

information about the proper proteome therefore it is necessary to develop an optimised protocol for each plant separately by trial and error (Newton et al. 2004).

c. Protein resolubilisation

After cell disruption and removal of interference substances, the protein molecules must be in a state of denaturation and reduction to disrupt intra-molecular and inter-molecular interaction and solubilised while staying at the inherent charged properties (Görg et al. 2004). TCA precipitation of proteins are unable to solubilize completely, thus a prolonged incubation of protein precipitation in the lysis buffer containing chaotropic (urea, thiourea), detergents (NP-40 or CHAPS), a reducing agent (DTT or 2-ME) and carrier ampholytes with shaking is able to assist re-solubilisation (Görg et al. 2004; Martínez-maqueda et al. 2013) (Table 2.5). For examples, Vâlcu & Schlink (2006) recommended a lysis buffer that contained DTT for leaves and needles and combination of tris (2-carboxyethyl) phosphine and DTT for roots for optimal extraction and resolubilisation. Isaacson et al. (2006) and Kumar et al. (2017) typically used a lysis buffer for plant protein samples in 2DE analysis containing 7M urea, 2M thiourea, 20 mM DTT, 2-4% CHAPS, and 2% carrier ampholytes. It is recommended that the lysis buffer selection is mostly based on empirical processes rather than identifying a definite rule for all samples; and the protein extract is in need of being clarified by carefully centrifugation to remove insoluble materials (Wu et al. 2014)

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

Buffer type	Characteristics	Example	Reference
Chaotropic agents	Destroy hydrogen and hydrophobic interactions, prevent protein aggregation.	Urea, a mixture of urea/thiourea.	Alam et al. (2013) De Filippis & Magel (2012) Kumar et al. (2017)
Amphoteric detergents	Dissolve proteins, prevent hydrophobic interactions	NP-40	Chen et al. (2011)
Reducing agents	Reduce disulfide bridges	DTT, 2-ME	Vâlcu & Schlink (2006) Isaacson et al. (2006)
Ampholytes	Increase solubility of proteins, decrease protein interactions,	Inositol phosphate glycan	De Filippis & Magel (2012)

d. Protein pre-fractionation/enrichment

Due to the limitation of analytical methods and the unknown complexity of the biological samples (e.g. wide range of protein concentration), it is desirable before analysis to reduce its complexity by pre-fractionation, and enrichment for proteins of interest (Bodzon-Kulakowska et al. 2007). The purpose of pre-fractionation is to obtain distinguishable

fractions containing restricted numbers of molecules, which can be conducted through various approaches including precipitation, centrifugation, filtration, and velocity or equilibrium sedimentation. As a result, the concentration of the proteins of interest will increase, prevent protein loss, especially low-abundant proteins which are believed to carry valuable diagnostic information and responsible for important processes in cells (Bodzon-Kulakowska et al. 2007).

e. Protein fractionation and detection

There have been several methods for protein samples preparation: electrophoresis, capillary electrophoresis, high-performance liquid chromatography, flow-field-flow fractionation, size by size exclusion chromatography (SEC) (Vensel et al. 2014) and lab-on-a-chip technology (Goetz et al. 2004). Each method has its own merits which, however, unable to cover all aspect of protein separation process as proteins possess high complexity in term of concentration and structure. Over past decades, electrophoresis in combination with other methods has been emerging approach for proteins separation and detection (Bodzon-Kulakowska et al. 2007; Anguraj Vadivel 2015; Lindemann et al. 2017).

The electrophoresis method based on using polyacrylamide gel electrophoresis in SDS-PAGE is implemented in the electric field filled with a solution containing a protein that has a net positive or negative charge, which migrates the protein at a rate, depended on its net charge, size and shape. While one-dimensional separation is often used as a pre-fractionation technique, two-dimensional gel electrophoresis (2-DE) method is now one of the most commonly applied techniques for studying proteome as it can increase the resolution power for highly complex samples (Westermeier 2014; Lindemann et al. 2017). The cooperation of isoelectric focusing (IEF) and SDS-PAGE is the most widely used separation tool in proteomic analysis (Pomastowski & Buszewski 2014; Ahmad & Ahmad 2014). In 2 DE, proteins are separated according to pI in the first dimensional IEF and secondly by molecular weight (Mt) in the second dimension SDS-PAGE (Rabilloud et al. 2010; Anguraj Vadivel 2015).

In the first dimension, proteins are subjected to separation by IEF depending on the presence of particular functional groups of the amino acids from which they are built, prosthetic groups or post-translational modifications (PTMs) (Mann & Jensen 2003), proteins have a specific electric charge. Base on the correlation between the pH of the medium and overall electric charge, and by manipulating the ambient pH, at a specific pH gradient, proteins migrate to a place where their electric charge equals zero – the so-called isoelectric point (pI) (Pomastowski & Buszewski 2014). There are two methods to create a pH gradient: using gradient gel with free ampholytes and using immobilized pH gradient (IPG). Due to the reduction of reproducibility of the first method, the use of IPG strips, which create a stable pH gradient, can improve the sensitivity and the reproducibility of the IEF. IPG strips are designed and selected at different lengths (7-30cm) and different pH ranges: wide (3-10), medium (3-6, 4-7, 7-11), narrow (6.3 – 8.3, 4.5-5.5) and ultra-narrow (4.9-5.3) for a different type of the biological materials. For

example, while a wide range of the pH is employed to illustrate the total proteins, the ultra-narrow pH gradient is employed to identify particular proteins and to determine their PTMs (Vadivel 2015).

In the second dimension, the IPGs are equilibrated by equilibration buffer and placed on the top of a polyacrylamide gel. SDS is an anionic detergent with a net negative charge which binds to most soluble protein molecules in aqueous solution over a wide pH range is added. The SDS terminates most of the complex structure of proteins from secondary to quaternary level, and reduces protein disulphide bridges (O'Farrell 1975). Also, the SDS confers to the polypeptide a negative charge which is utilised to separate the protein in an electrical field within PAGE (O'Farrell 1975). Finally, the separation is carried out vertically and horizontally.

The advantages of 2-DE are enormous with the detection limit (for visualisation of the protein when dyed) less than one ng of protein. Also, it can create protein patterns (profiles) that can be analysed by image analysis software. The main disadvantage of 2-DE is that it is fairly expensive technique, time-consuming, sometimes ends up in streaky 2D patterns in the 2-D gel due to the insufficient solubility of particular proteins caused by overloading, protein interactions with contaminations and dismissed solubility near the pI or poor background (Ahmad & Ahmad 2014). Another drawback of the 2-DE is the reduction of the reproducibility of samples separation. To solve this problem, a method namely two-dimensional difference gel electrophoresis (2-DIGE) which can separate two samples on the same gel simultaneously has been developed (Marouga et al. 2005; Westermeier 2014). This method enables elimination of gel variability, and can provide distinguishable information in migration due to pI or Ms and widely applied in clinical identification of potential biomarkers, PTMs or changes in the proteome expression being analyzed (Marouga, David & Hawkins 2005; Martínez-Esteso, Sellés-Marchart, et al. 2011; Corchete & Bru 2013).

After being separated, proteins are detected using various staining methods such as fluorescent labelling, Coomassie Brilliant Blue (CBB), negatively charged metal salts (zinc, copper or silver). Of these methods, staining with negatively-charged silver is not compatible with mass spectrometry (Pomastowski & Buszewski 2014).

f. Protein analysis

Although the 2-DE has several disadvantages, it has been the most commonly used method due to its possibility to couple it with a number of analytical techniques such as capillary electrophoresis (CE), flow-field-flow fractionation (4F) and high-performance liquid chromatography (HPLC) for protein fractionation, and Edman degradation and mass spectrometry (including electrospray ionization (ESI), matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) and surface-enhanced laser desorption/ionization (SELDI)) and western blotting (WB) for proteins identification (Pomastowski & Buszewski 2014; Westermeier 2014; Jorrín-Novo et al. 2015).

Recently, such improvements in proteomics as high-resolution protein separation, MS software and hardware, and advanced bioinformatics technology have led to an increasing application for proteomics in all biological fields (Schubert et al. 2017). In the field of profiling proteins, 2-DE combined with MS has proved to be the most powerful method (Wang et al. 2008). Of recent combinations, the use of 2-DE with MALDI-TOF-MS is the most common tool for protein analysis (Angel et al. 2012; Agrawal et al. 2013; Jorrín-Novo et al. 2015; Schubert et al. 2017). The MALDI system utilises radiation of a nanosecond laser to cause desorption and ionisation of analytes. Since it is typically coupled with TOF analyser, separation of ions is based on the different velocities of ions with the same kinetic energy and different masses. This method makes polypeptides accessible to mass spectrometric analysis (Domon & Aebersold 2006). The major advantage of MALDI-MS is that it can be used to analyse high molecular weight, high throughput, sensitivity, good mass accuracy, ease of operation and automation (Pomastowski & Buszewski 2014).

g. Bioinformatics

Bioinformatics analysis of protein has become a fundamental part of molecular biology research, especially for information obtained from MS, which can be analysed by algorithms and dedicated bioinformatics databases. As a result, data on the molecular weight and the amino-acid sequences of the proteins will be identified. Several methods have been developed such as De novo sequencing, peptide mass fingerprinting (PMF), database searching, and automation and robotics. *De novo* sequencing involves determining the unknown peptide sequence. Therefore it is used for mapping amino acids from a peptide. PMF is also used for peptide mapping or peptide-mass mapping via on-line PMF matching with proteins from a database. Therefore it is suitable for identifying protein orientating from organisms whose genomes have been fully described (De Filippis 2013). Database searching provides a helpful tool as protein sequence is available. For example, there have been a number of easy-to-use web services which have been developed, including the NCBI resources (http://www.ncbi.nlm.nih.gov), the SIB Bioinformatics resource portal (ExPASy; http://www.expasy.org) (De Filippis & Magel 2012; De Filippis 2013), the EMBL – EBI Bioinformatics web services (http://www.ebi.ac.uk/services), the Protein analysis toolkit (PAT; http://pat.cbs.cnrs.fr), the PredictProtein server (https://www.predictprotein.org), and the CBS prediction servers (http://www.cbs.dtu.dk/services) (Alva et al. 2016). These require effective algorithms in computer programs such as Mascot, Sequest, Tandem, MS-Blast and Peak. Recently, the 2-DE-coupling with MALDI-TOF-MS involves automation, and the deposition of a prepared protein on a MALDI target can be performed by robots (De Filippis 2013). Although a method for automation and robotics has been developed, the full procedure is lacking in both experiments and software (Kaczmarek et al. 2002).

As the fully sequenced genomes of non-model plants such as medicinal plants species have not been available in public databases, the identification of the specific proteins related to secondary metabolism is significantly challenging (Martínez-Esteso et al. 2015; Rai, Saito & Yamazaki 2017). In order to overcome the problems, different approaches

have been adopted in recent studies, which use sequence homology to *Arabidopsis thaliana* and other plant species (Jacobs et al. 2005; Cheng & Yuan 2006; Bhattacharyya et al. 2012; Nagappan et al. 2012) or EST database (Desgagné-Penix et al. 2010; Schilmiller et al. 2010; Champagne et al. 2012). Since sequence annotation from model plant species usually contains a very low diversity of secondary metabolites, the application of specific EST databases such as BLAST, BLAST2GO tool have been used successfully in the case of annotating proteomic experiments in grapevine (Martínez-Esteso, Casado-Vela, et al. 2011).

2.4.6 Conclusion

Celastrus hindsii Benth has been found to contain variety secondary metabolites and vitamin including sesquiterpene, triterpenes, alkaloids, flavonoids, vitamin C and vitamin E through a range of recent phytochemical investigations by Hu et al. (2014); Thuy, Cuong & Sung (2007); Ly, Shimoyamada & Yamauchi (2006); Yao-Haur Kuo, Chen & Kuo (1995). These compounds are shown to have some health effects to reduce the incidence of cancer and cardiovascular diseases. With the abundance of diversely specialised metabolites, the medicinal plant *C. hindsii* has been considered as a promising candidate for the pharmaceutical industry, which requires a large-scale medicinal plant material production. Advanced technologies have been brought into the field to increase the yield of the pharmaceuticals of the interest, however, the production of secondary plant metabolites is often restricted to a species, activated only during a particular growth or developmental stage, or under specific seasonal, stress or nutrient availability conditions, and regulated by molecular mechanism. Recent studies conducted on medicinal plant *C. hindsii* have revealed valuable phytochemical properties. However, the knowledge of secondary metabolite pathways, in fact, is limited. Out of molecular study disciplines, proteomics has brought further insight into plant secondary metabolism, which is essential for altering biosynthesis pathway with the purpose of increasing the level of desired compounds in the targeted plant. Proteomics has broadened our understanding about the biosynthesis pathway of secondary metabolites and their regulation. However their application is still challenging (Martínez-Esteso et al. 2015; Rai, Saito & Yamazaki 2017), especially for non- model plant such as *C. hindsii*. Therefore, the selection of appropriate methods from sample preparation to protein detection and identification, and protein analysis for implementation is crucial.

CHAPTER 3

STUDY SITES AND SPECIES

3.1 An overview of *Cycas fugax*

3.1.1 Introduction

Cycads represent a very primitive group of vascular plants that have been in existence for more than 200 million years (Hendricks 1987). Nowadays, they have been declined dramatically and the recent distribution only in tropical and subtropical regions. According to the survey of Osborne (1995), the majority of wild cycad population are either threatened, critically endangered, or near extinction. Thus, all cycads species have been listed in the Convention on International Trade in Endangered Species of Wild Fauna and Flora. Generally, Vietnam has the greatest diversity at the sectional level and the highest number of species of any country except Australia (Hill et al. 2004). The genus of Cycas in Vietnam includes 27 species, and nine of them described as new, one species is listed critically endangered (*Cycas fugax*), three are considered endangered (*C. aculeate, C. hoabinhensis and C. multipinnata*) while the others are either vulnerable or near threatened (Osborne et al. 2007).

C. fugax is one of the rarest cycad species which is assessed as critically endangered (CR) (possibly extinct) in 2010 IUCN Red List in Vietnam (Osborne et al. 2010). The species is known only from Phu Tho province in the north of Vietnam and occurs at 200 m above sea level (Osborne et al. 2007). It is estimated that more than 80% of *C. fugax* population has been reduced due to natural habitat destruction by agriculture encroachment and urbanisation (Osborne et al. 2010). Consequently, extent of occurrence is also extremely small, and if populations survive, they are extremely fragmented and continuously declining. This species is urgently in need of both close study to evaluate the conservation status, and protective measures to save it from extinction. Up to date, there have not been any studies on *C. fugax* with regards to investigating genetic diversity recorded in the literature.

3.1.2 Taxonomy

Cycas fugax K.D.Hill, T.H.Nguyên & P.K.Lôc, belongs to the genus *Cycas*, family Cycadaceae. From the Latin *fugax*, fleeting or ephemeral, referring to the near extinction of this species before it was recognised as a botanical entity. The species was discovered and recognised as a distinct species only in late 1990's. The species falls within the 3-species group that Hill (2004) refer to as the *C. simplicipima* complex, also including *C. collina.* The common characteristics of this group are: a subterranean habit; reduced megasporophylls that lack a distinct apical spine. However, *C. fugax* is distinguished from the other two close species by its few very long leaves (1-3 leaves), widely spaced leaflets with very narrow tapering bases, small male cones with microsporophylls that have a terminal spine, and megasporophylls with elongated teeth.

3.1.3 Distribution and conservation status

C. fugax is endemic in Vietnam, as only know from Phu Tho province in north-eastern Vietnam, naturally distributed at about 200 m in altitude. The first specimens were collected in early 2000 at Tram Than commune, Phu Ninh district, Phu Tho province and firstly described by (Hill et al. 2004). They originally distributed in the evergreen forest on low hills. However, recently the habitats have been almost destroyed for agriculture or plantation such as tea crops, *Eucalyptus* and *Acacia* plantations.

Because of habitat loss due to agriculture and plantation expansion, there have been only a few representatives in the wild. Several plants are found in home gardens in the villages near the species' original habitat, in which they are usually planted as a single specimen or in clusters of two (male and female). In fact, the gender of each had not been assured by the time of being cultivated by villagers. *C. fugax* is the most threaten cycads species in Vietnam with total population size from 250 - 2000 occurring over an area of 200 km^2 (Osborne 2010). However, this number may be over-estimated. *C. fugax* is only Vietnam' cycad which is listed as critically endangered (CR) in IUCN Red List (Osborne et al. 2010; Osborne et al. 2007).

3.1.4 Morphological characteristics

C. fugax has a subterranean flask-shaped stem, about 18 cm long, 15 cm in diameter at the base and 10 cm at the apex, which may be larger and branched in older specimens. Stem hold 1-3 leaves in the crown. Leaves are bright to deep green, moderately glossy, with an orange tomentum which is lost as the leaf expands, and from 280 – 380 cm long and flat (not keeled) in section. There are 80-100 leaflets, and the rachis ends consistently with a pair of leaflets. The basal leaflets are 12 – 30 cm long and are abruptly replaced by petiolar spines. The median leaflets are simple, strongly discolourous, 40 – 50cm long, 18 – 27 mm wide, angled forward (inserted at $70\text{-}85^0$ to the rachis), decurrent for 5-10 mm, narrowed to 2-3 mm at the base, 30-40mm apart on the rachis, flat in section, with margins flat or wavy, with apex softly pointed and with midrib raised above but bot below. The petiole is 150-220 cm long (50 – 60% of total leaf length), glabrous and spine scent for 90-100% of its length (Osborne et al. 2007).

Male cones are solitary, spindle-shaped, cream in colour, 10-12 cm long and 2.5-4 cm in diameter. Microsporophyll has a soft blade that is not dorsoventrally thickened, 12 – 14 mm long, 7.5 – 10 mm wide. Megasporophylls is 20-25 cm long, brown-tomentose and has 2-4 glabrous ovules. The megasporophylls blade is ovate, 110 – 130 mm long by 35-40 mm wide, deeply pectinate with 10-14 soft spines which are 50-90 mm long and 3 mm wide, and the apical spines is not distinct from the lateral spines. Seed is ovoid, 25 – 27 mm long, and 18 – 21 mm wide. The sarcotesta is yellow, not powdery, and non-fibrous. The sclerotesta is warty, and there is no internal spongy layer (Osborne et al. 2007).

Figure 3.1 Vegetative and reproductive organs of *C. fugax* (A: Only two leaves on crown; B: Brown-tomentose megasporophylls with deeply pectinate; C: Ovoid seeds)

3.2 An overview of *Celastrus hindsii* research

3.2.1 Introduction

The species *Celastrus hindsii* Benth belong to the genus *Celastrus* which include species of aromatic herbaceous perennials, almost exclusively native to Asia. Research on this species and the genus *Celastrus* has covered several scientific disciplines but has been disproportionately focussing on a few species such as *C. paniculatus, C. angulatus, C. orbiculatus, C. hinsdii* and mostly conducted in investigating chemical and phytochemical constituents and their bioactivities. *C. hindsii* has been the main foci of studies on the uses, biochemical and cultivation. Other species, such as *C. obovatifolius, C. yuloensis, C. caducidentatus* are generally only studied in taxonomic disciplines as the first time described species (Mu et al. 2012; Ao et al. 2012; Liang et al. 2016). In this section, a critical review including much of the current and past work on the genus *Celastrus* with the main focus on the research species *C. hindsii* to present an overview

that briefly describes the morphology, taxonomy, biochemistry, uses and cultivation aspects of this increasingly popular species.

The genus *Celastrus* Linnaeus belongs to the family Celastraceae and is comprised of about 35 species, distributed in the subtropics and tropics, with a few representatives in temperate areas around the world, but higher density in East Asia, Oceania, both North America and Madagascar with present centre of distribution lying in Southeast Asia (Liang et al. 2016). In Vietnam, the genus *Celastrus* was known to have eight species: *C. hindsii, C. genatus, C. hookeri, C monospermus, C. orbiculatus, C. annamensis, C. tylosus, C. paniculatus* (Ban 2003). The genus is a deciduous or evergreen woody vine and featured by alternate simple leaves, axillary or terminal cymes with mostly 5-merous pale green bisexual or unisexual flowers, 3- to 6-seeded capsules and seeds (≤5 mm in length) covered by a distinct red or orange fleshy aril (Mu et al. 2012). The genus also can be distinguished by its typically scandent shrubs, a distinct articulation on the stalk, consistent 3-valved fruits and complete cup-shaped aril. Due to overlapping morphologies inter and intraspecies, mainly variable leaves and inconspicuous flowers, several species have been defined in the conflict in the literature (Zhixiang & Funston 2008). Leaves are elliptic to oblong, or broadly ovate to orbicular (Liang et al. 2016). Although recent molecular studies investigating the phylogenetic relationships within the genus *Celastrus* have been implemented (Simmons et al. 2012; Simmons et al. 2008), the further extensive molecular phylogenetic investigation is still required to understand the relationships within the genus and confirm existed subgeneric classifications (Liang et al. 2016).

The plants of *Celastrus* have been used as natural insecticides (Lusby 1988), and also as important folk medicine to treat fever, chill, joint pain, edema, rheumatoid arthritis, and bacterial infection in Asia for long time as they produce a large spectrum of structurally and biogenetically diverse secondary metabolites, such as sesquiterpenes, alkaloids, triterpenes, diterpenes, and flavonoids (Zhang et al. 2009). For example, *C.hinsdii* contains some triterpenes exhibited antitumor and anti-HIV activities (Huang et al. 2000; Sung et al. 2008; Kuo & Kuo 1997).

In China, plants of the genus *Celastrus* have been used as traditional herbal medicine treat several of diseases and disorders in Chinese folk medicine. For instance, *C. orbiculatus* has been used as folk remedy for rheumatoid arthritis, low back pain, muscles pain, toothache, amenorrhea, dysentery, bruises, snack bites and bacterial infection (Li et al. 2012). In Himalayan folklore medicines, *C. paniculatus* is reported in the treatment of haemorrhoids, piles, gout, rheumatism, cold, dysentery, diarrhoea, leprosy, snake bite, wounds (Younus 2015) and to refresh mental activity (Ning et al. 2015). People in Eastern Tibet of Chian have used the roots of C. rugosus for treatment of rheumatism, strain, and measles in children (Ying et al. 2017). The term and the roots of *C. oblanceifolius* have been used as medicine to cure injuries caused by snakes, while the bark and leaves have been used as a pesticide (Wang & Wang 2016).

In some South East Asia countries like Indonesia, Thai Lan, Philippines, the most common part of *C. paniculatus* such as aril, seeds or seed oil have been used to cure pimples, rheumatism, muscle pain and paralysis. In India traditional system of medicine, *C. paniculatus* is used as appetiser, laxative, emetic, aphrodisiac and used for the treatment of cough and leukoderma (Younus 2015), and for alleviating cognitive disturbances (Ning et al. 2015).

In Vietnam, the seeds of *C. paniculatus* have been used to treat some specific diseases such as the use of to treat rheumatism and leprosy; fatty oil extracted from seeds to treat back pain, hair loss, digestive disorders. Leaves are used as detoxifying drugs and stimulants (Muoi et al. 2009). *C. hindsii* have been used for generations in Northern Vietnam, particularly by Muong people for diseased treatment relating to ulcers, tumours and inflammation. Besides, leaves of *C.hindsii* have been used for the manufacturing of tea products as a healthy drink (Ly et al. 2006).

An increasing number of publications indicate that *C. hindsii* is a candidate for large-scale cultivation as a source of aromatic and medicinal plant for its essential secondary metabolites such as flavonoids (Ly et al. 2006), alkaloids (Yao-Haur Kuo et al. 1995; Huang et al. 2000; Hu et al. 2014), as well as their bioactivities which are important to the species physiology and human being. We will emphasise studies focusing on morphology, genetic diversity, metabolism and their pathway to support future conservation and cultivation in next sections and chapters.

3.2.2 Taxonomy and morphology

C. hindsii Benth belongs to the genus *Celastrus* Linnaeus, the family Celastraceae. In the cluster analysis (Figure 3.2), *C. hindsii* is placed in the same clade as *C. virens* which shows the closest morphological similarity (Hou 1955). They also show a close relationship with *C. subspicatus* which is distributed in the Ocean. However, in a later study assessing the phylogeny of *Celastrus* L., based on sequences of two nuclear (ETS or ITS) and three plastid (*psbA-trnH, rpl16 and trnL-F*), Mu, Zhao & Zhang (2012) reported that *C. hindsii*, a synonym of *C. tonkiensis* is nested within the cluster of *C. monospermus*. *C. tonkinensis* is still retained because of its broadly elliptic (vs narrowly elliptic) leaves, transverse (vs. smooth) on valves (Muoi et al. 2009; Zhixiang & Funston 2008). However, these differences are not constant, more anatomical and phylogenetic works are needed for elucidation of the relationship between them (Mu et al. 2012). While *C.tonkinensis* is endemic to the Karst area of southwest China and Vietnam, *C. hindsii* has a wider range of distribution in forests, thickets and mountainous regions with altitude from 200-2500 m in China, Northern Vietnam, India, Malaysia and Myanmar.

In Vietnam, *C. hindsii* is wildly distributed in the Karst areas of Hoa Binh, Hai Phong, Ha Noi and Ninh Binh province, and recently cultivated in many other Northern provinces such as Phu Tho, Tuyen Quang, Quang Ninh, Bac Ninh, Bac Giang, Ha Nam, and Thu Thien Hue, Gia Lai in the central of the country (Muoi et al. 2009).

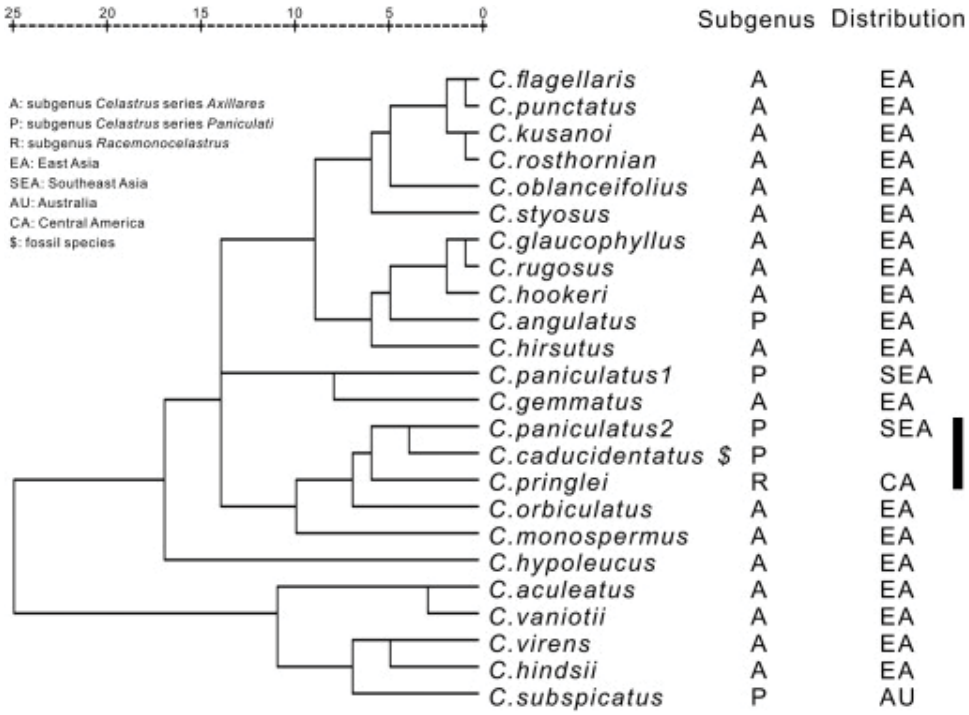


Figure 3.2 Subgeneric classification of *Celastrus* and distribution and nearest living relative of *C. hindsii* in leaf morphology shown by cluster analysis (between-groups linkage) (Hou 1955).

According to the description of (Zhixiang & Funston 2008), *C. hindsii* Benth is an evergreen twining shrub. Branchlets are terete, grey or purple, with very sparse lenticels. Axillary buds are ovate-triangular, 1-1.5mm long. The petiole is 6-10 mm long. Leaves blade are narrowly rectangular-elliptic, or narrowly ovate-elliptic to elliptic-oblanceolate, 7-14 × 3-6 cm, papery or leathery, often grey-green when dry, with the cuneate or rounded base, margin sparsely serrate, apex acute, caudate-acuminate to acuminate. Venation consists of 5-7 of secondary veins, veinlets between secondary veins parallel, prominent. Reproductive organs have terminal and axillary thyres, 5-14 cm long, axillary inflorescences 1-3 flowered with pedicels 4-5mm long, joined at the upper middle of the peduncle. Flowers are light green with approximately semicircular sepals and rectangular petals, margin ciliate. Flower disk is cupuliform, membranous, shallowly lobed, lobe triangular. Stamens are inserted into the margin of disk, with subulate filament and ovate-orbicular anther. Pistils look like ampulllaceous shape with approximately globose ovary; styles ca. 1mm; stigma slightly 3-lobed. Fruits are approximately globose, 7-9x 6.5-8.5mm with style persistent on young fruits, up to 1.5mm and valves are slightly rugose. Seeds are broadly elliptic to globose, 5-8 mm long, with orange aril. Flowers bloom from May to July and fruits formation is from July to October (Muoi et al. 2009).

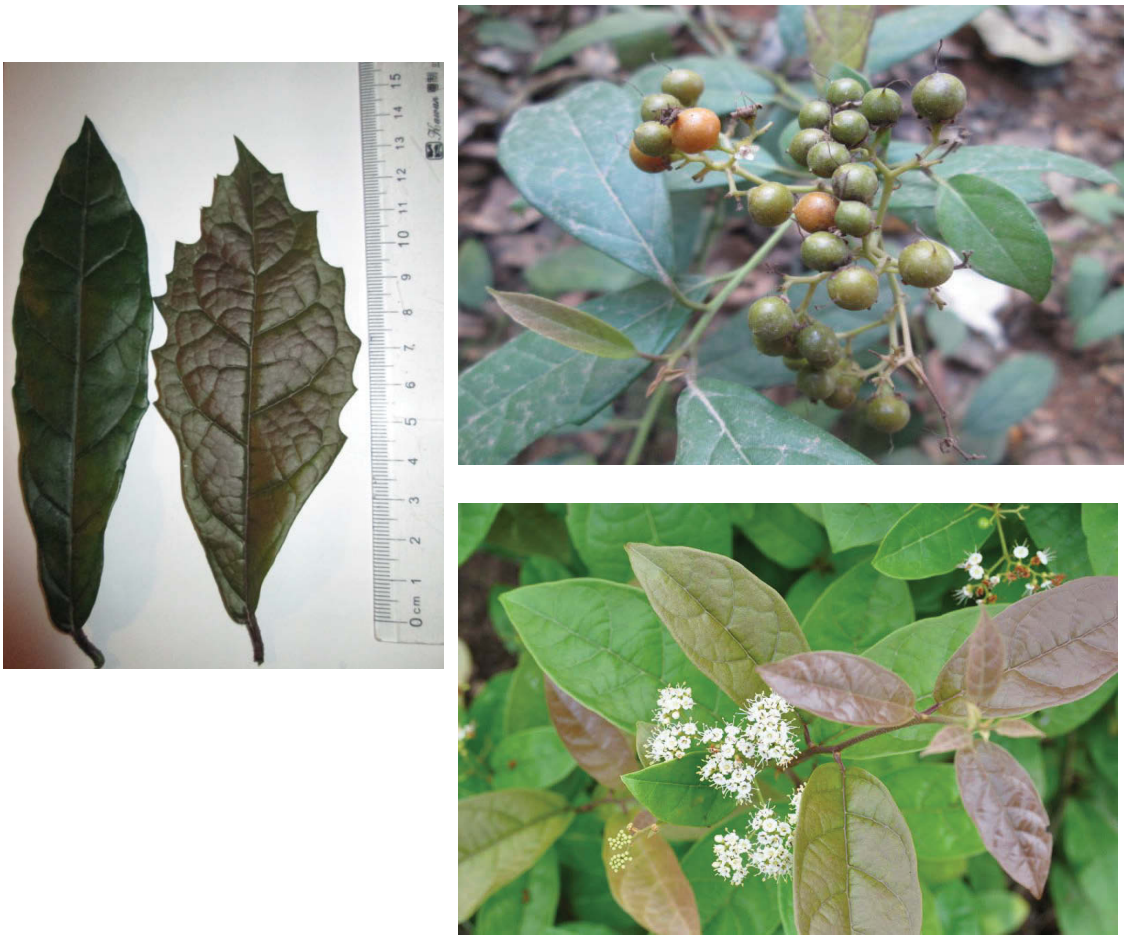


Figure 3.3 Vegetative and reproductive organs of *C. hindsii* (A. Mature and young leaf; B. Globose fruits with style persistent on young fruits; C. Thyres)

3.2.3 Biological activity study

Over last decades, a large number of secondary metabolites exhibiting a wide range of bioactivity have been extracted from *C. hindsii*. In addition to numerous terpenoids, including a diverse array of sesquiterpenoids, various bioactive alkaloid and flavonoids have also been isolated (Hu et al. 2014). The most pervasive and characteristic metabolites of *C. hindsii*, however, are a large family of unusually highly oxygenated sesquiterpenoids and it is this class of compounds that has attracted the most interest.

In particular, celahinine A, a sesquiterpene pyridine alkaloid and the related known polyester A which shows potent cytotoxicity against Hepa-2 (hepatoma), Hela (cervix carcinoma), COLO- 205 (colon carcinoma), and KB (nasopharynx carcinoma) cells *in vitro* was isolated from dried stems of *C. hindsii* by (Yao-Haur Kuo et al. 1995). In their continuous research two years afterwards, four new triterpene compounds, celasdin-A, celasdin-C, anti-AIDS celasdin-B and cytotoxic maytenfolone-A were also isolated from

the dried stems of *C. hindsii*. Of those four, maytenfolone-A demonstrated cytotoxicity against hepatoma and nasopharynx carcinoma and celasdin-B was found to exhibit anti-HIV replication activity in H9 lymphocyte cells. In another similar study, Hui 2000 have isolated a new sesquiterpene, celadin D and other six related derivatives in which ermarginatine-E exhibited cytotoxicity against KB and COLO-205 with ED_{50}=1.7µg/ml and ED_{50}=4.1µg/ml, respectively, whereas the ED_{50} for other compounds all exceeded 10µg/ml.

Antioxidant compounds such as vitamin C, vitamin E, and flavonoid polyphenols which may have some health effects to reduce the incidence of cancer and cardiovascular diseases have been isolated from dried leaves of *C.hindsii* collected in Ha Tay province, Vietnam. Eight phenolic compounds including rutin, kaempferol 3 rutinoside, rosmarinic acid, lithospermic acid, lithospermic acid B and three novel oligomers of rosmarinic acid, a dimer and two trimers were finally obtained by reversed-phase high-performance liquid chromatography. These phenolic compounds were shown to have antioxidant activities against the autoxidation of methyl linoleate in bulk phase and the radical-initiated peroxidation of soybean phosphatidylcholine in liposomes (Ly et al. 2006). In other isolations from *C. hinsdii* leaves also collected in Vietnam, Thuy, Cuong & Sung (2007) reported the structure elucidation of three triterpenes, 3-friedelanol, 3-friedelanone and canophyllol which were determined by MS and NMR spectroscopic data.

Recently, Hu et al. (2013) isolated three new diphenylpropanes named Hindsiipropane A, B, C, together with one known arypropyl quuinone Griffithane D which is firstly obtained in the genus *Celastrus*. All these isolated compound show modest cytotoxicity against four human tumour cell lines (A549, HCT116, MDA-MB-231 and BEL7404) with IC_{50} values in the range of 10.95-62.19 µg/ml. In the later study by Hu et al. (2014), a new macrocyclic lactone named Hindsiilactone A, a new 5,8-quinoglavan named Hindsiiquinoflavan B, and three known compounds combretastatin D-2, combretastatin D-3 and isocorniculatolide A were isolated from the stems of *C.hindsii.* The first four compounds are confirmed to exhibited moderate or weak cytotoxicity toward the human small-cell lung cancer (NCI-H187), colon cancer (HCT116) cell lines with IC_{50} values ranging from 14.9 to 36.8 µg/ml. The fifth compound showed moderate or weak cytotoxicity toward the breast cancer (BC-1) and liver cancer (HuH7) cell lines with IC_{50} values of 19.8 and 21.2 µg/ml.

All above studies have revealed a variety of chemical components from *C. hindsii*, including sesquiterpene, triterpenes, alkaloids and flavonoids as well as their bioactivities. However, the findings from these studies withdrew from studies conducted on a random selection of plant materials and mostly from the stems and leaves. While *C. hindsii* has been widely distributed both in the wild and by cultivation with different climates and biotic and abiotic conditions which contribute greatly to the variability and abundance of secondary metabolites, the phytochemical investigation which conducted on plant material collected from a single location may not be representative for the species. Also, all previous studies on *C. hindsii* focus on chemical structure and

bioactivity of secondary metabolites, but none of them investigated the genome or proteome of the species to understand their genetic diversity and metabolites biosynthesis further. This information is necessary to set up priority in conservation and breeding programs.

3.2.4 Breeding study and cultivation state in Vietnam

Studies on breeding of *C. hindsii* have rarely been conducted over the world and in Vietnam, and almost all research findings have not been published or primarily published on unscientific magazine both paper or electronic versions providing information about dosage in combination of drugs and medications. In 1987, *C. hindsii* was firstly studied by medical experts in Military Medical University, Vietnam. According to their unpublished report in 1999, extract from *C. hindsii* contain valuable active secondary metabolite such as alkaloid, triterpenoid, flavonoid which can inhibit the cancer cell development. Although, *C. hindsii* has been used as a medicinal plant for cancer disease by Muong minority group in the north of Vietnam for generations. As soon as the cancer treatment potential of *C. hindsii* has been made public, the plant has been overexploited in the wild by local people for domestic use as well as commercial purposes. While the wild source of *C. hindsii* has been almost exhausted, the increasing demand for plant material used for folk medicine increasing has led to the development of cultivations in several provinces with different purposes and scales ranging from home garden to intensive nursery for domestic use or trading as a good.

To meet the increasing demand of the market, several breeding studies have been conducted such as clonal propagation and tissue culture. Loan (2012) used stimulants NAA at a concentration of 1000ppm which generated the best result of survival rate at 75.56%. *C. hindsii* was also successfully propagated through tissue culture in which Nodal segments were cultured to induce multi-shoots on Murashige and Skoog's (MS) base medium supplemented with 5 mg/l 6-benzylaminopurine (BAP), 20 g/l sucrose, and 7 g/l agar. *In vitro* raised shoots were rooted on the ½ MS medium supplemented with 0.5 mg/l indole-3-butyric acid (IBA), 0.5 mg/l BAP, 20 g/l sucrose, 7 g/l agar, and 1 g/l activated carbon, which resulted in 81.91% of total samples inducing multi-shoots and the highest rooting percentage (94.8%), respectively (Nam et al. 2013).

3.3 The study sites

In order to assess genetic diversity of *C. hindsii*, the study conducted plant material collected from various sites in four provinces in the north of Vietnam including Hanoi, Hai Phong, Hoan Binh, Phu Tho (Figure 3.4, Table 3.2). *C. fugax*, however, as stated in literature only distributes in Tram Than commune, Phu Tho Province (Figure 3.4, Table 3.1). The details of collection sites will be presented in following sections.

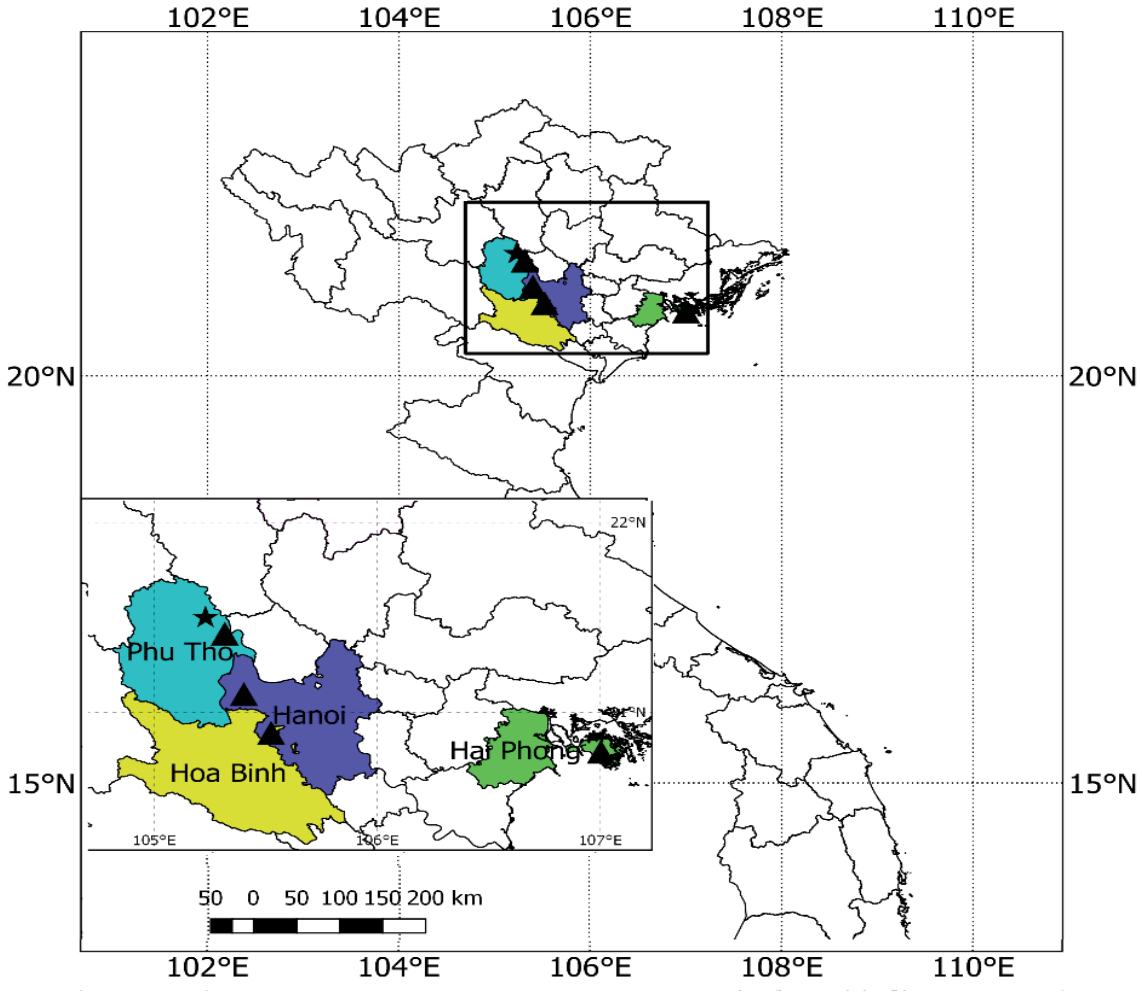


Figure 3.4 Study locations of *C. fugax* (only in Phu Tho province) and *C. hindsii* (in four provinces in the north of Vietnam: Phu Tho, Hoa Binh, Ha Noi and Hai Phong). Sitemap for a sampling of *C. fugax* (★) and *C. hindsii* (▲).

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.

Code	Location names	Coordinate location (from – to)		Distance to the nearest site (km)	Ecosystem, topography
LV1	Lam village 1	21^0 30’45.6620”N	21^0 30’45.6620”N	0.65 km to NF	Home garden, lowland
		105^0 13’1.5870”E	105^0 13’1.5870”E		
LV2	Lam village 2	21^0 29’56.7121”N	21^0 30’8.971”N	2 km to LV1	Home garden, lowland
		105^0 13’40.875”E	105^0 13’48.606”E		

NF	Natural forest	21^0 31’2.701”N 105^0 12’53.805”E	21^0 31’9.75”N 105^0 13’1.343”E	0.65 km to LV1	Natural forest, steep hill
TV	Than village	21^0 26’44.7064”N 105^0 14’57.432”E	21^0 26’44.7064”N 105^0 14’57.432”E	6.5 km to LV2	Home garden, steep hill

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

Location	Coordinate location (from – to)		Distance to the nearest site (km)	Ecosystem, Topography
Phu Tho (PT)	21°20'24.3"N 105^0 18’24.4721”E	21^0 24’38.6221”N 105°22'48.9"E	32 km to HN	Home garden, lowland
Ha Noi (HN)	20^0 48’9.3177”N $105^0$21’49.6249” E	$21^0$6’24.4650”N 105^0 38’43.616”E	32 km to PT	The buffer zone, home garden, lowland
Hoa Binh (HB)	20°39'53.0"N 105^0 23’19.3060”E	20^0 53’59.8983”N 105°40'12.1"E	33.75 km to HN	Intensive farming, home garden, Lowland
Hai Phong (HP)	20^0 47’37.321”N 106^0 58’54.25”E	20^0 50’18.953”N 107^0 3’8.86”E	147.5 km HB	Natural forest, Limestone mountain

3.3.1 Ba Vi National Park, Ha Noi City

a. Topography and geology

Ba Vi National Park is generally located between $21^0$1’- $21^0$07’N and $105^0$18’- $105^0$25’E in the Ba Vi district, Hanoi capital, northern Vietnam, approximately 50 km north-west of Hanoi, with altitude ranging from 100 to 1296 m above sea level. The National Park was established in 1992, covers an area of 7377 ha, including a strictly protected area above the 400m elevation of 3000 ha surrounding three mountain peaks where there is still relatively pristine tropical montane and semi-evergreen forest at higher elevations (On et al. 2001; Phuc 2017). The mountain rises steeply out of a plain that rarely exceeds 30 m in elevation. Geology comprises mostly fine-textured sedimentary rocks (On et al. 2001).

The park area embraces all lands above 100 m. For the management purposes, three zones were designated in the park: the land from the 400m contour upwards was considered to

hold the richest biodiversity and was marked for strict protection; the 100m – 400m zone was identified as an ecological rehabilitation zone, the buffer zone (100m and below) of approximately 15,000 ha, where a range of non-park activities such growing crops, cattle grazing are permitted (Indochina. 2004).

b. Climate

Ba Vi National Park is located in the sub-climatic zone of North-Eastern Vietnam, which has a mild tropical climate, dominated by the summer monsoon. Winters are cool and relatively dry, summers hot and wet. The majority of annual rainfall occurs between June and October. The mean annual temperature is 22^{o} C, the average temperature of the hottest period is 39^{o}C (in July), and the lowest temperature is 0.6^{o}C (in January). The annual mean humidity is 83%. Total annual rainfall is 1,660 mm/year, 90% of which is received during the wet season months. Because of the mountainous topography, the climate at Ba Vi varies with altitude. Above 500 m, fog enshrouds the top of the mountain on most days (On et al. 2001; Indochina. 2004).

c. Hydrology

The Da River is the only large water body on the western side of the national park, and there are no large, permanent water bodies on the other sides. Streams in the national park are small, steep-sided and fast flowing. In the rainy season, the volume of water flowing through these small watercourses and over the surface of the ground is sometimes sufficient to cause landslides. In the dry season, however, many of the streams are dry (Indochina. 2004).

d. Vegetation

At Ba Vi national park, natural forest is mainly distributed at an elevation above 600 m with various of forest types such lowland evergreen forest, lower montane evergreen forest, lower montane mixed coniferous and broadleaf forest. Since 1998 beside a natural forest, the national park supported nearly 3,000 ha of plantation forest (Indochina. 2004; On et al. 2001).

According to the inventory of the French botanist, Balansa, at the end of the nineteenth century, 5,000 plant specimens were collected from the Ba Vi region. Anon (1991) revealed more than 800 species of vascular plants, occurring in 427 genera and 98 families. On et al. (2001) estimated the park' flora comprise 1500-200 higher plant species. Of those described are 13 local endemics, ten species recorded in the Vietnamese Red Data Book for rare and endangered plants. There are more than 300 local plant species collected and used as their medicinal value (The Asian Foundation 2012).

Since the establishment of the protected area at the site, exploitation of natural resources for fuelwood, timber, medicines and foods from local communities which happened legally before was prohibited. However, continuous exploitation activities still occur for both domestic used and commercial purposes and thus led the biodiversity of the national

park to a dramatic decline. Logging activities both by local people and forest enterprises, agricultural encroachment from the lowlands and shifting cultivation, fuelwood and medicinal plants collection, and forest fire have been responsible for the biodiversity loss and forest degradation. Medicinal plant collection was assessed as a major economic activity in the area since 250 tonnes of plant materials were extracted from the national park between 1997 – 1998 by 80% of population of the Dao ethnic group in Ba Vi commune (Indochina. 2004).

3.3.2 Cat Ba National Park, Hai Phong City

a. Topography and geology

The Cat Ba National Park is located between $20^0$42'- $20^0$54'N and $106^0$54'- $107^0$09'E in Cai Hai district, approximately 50km east of Hai Phong city in the northern section of the Tonkin Gulf. The archipelago consists of 366 offshore islands, and Cat Ba is the largest island. The park was established in 1986 by the Vietnamese government as its importance for nature conservation, covers an area of 15,200 ha comprising 9,800 ha islands and 5,400 ha of surrounding marine waters (MONRE 2015). Cat Ba national park is distinguished from other national parks in Vietnam as its diverse landscapes and ecosystems, including sand beaches, mangrove, freshwater swamp forest, tropical evergreen forests, small freshwater lakes and streams, and many coral reefs and seagrass beds (Indochina. 2004). In 2003, the Cat Ba archipelago was recognised as Vietnam's third biosphere reserve by the Standing Committee of the Man and the Biosphere Programme of UNESCO.

The karst limestone mountains rising abruptly from the sea are the main types of the landscape of the Cat Ba National Park. The topography is rugged and marked by steep outcrops and areas of bare rock. There are several high mountain peaks ranging from 230-331m with narrow valleys at between at 20-100m elevation. The beaches have developed along the south-eastern parts of the main island. Carboniferous-Permian limestone and dolomite is the main feature of the island geology, which is about 500-600m thick and well bedded. However, in some areas, vertical dipping bed rock occurs strongly (Indochina. 2004).
Although the limestones are the main geological characteristics, there are several types of soils have been formed and developed along with the decomposition processes in different areas on the islands. On the tops of the hills, the soil is being weathered by feralitic processes on clayey schist, while in wetlands which have been developed in mangrove forests, the alluvia are being deposited.

b. Climate

The climate in Cat Ba island is featured by the coastal tropical monsoon with wet and dry seasons. The wet (or rain season) lasts from May to September with a high frequency of typhoons and tropical rainstorms and the dry (or cold season) starts from November to March of following year. The average annual temperature ranges from 23^0 C to 28^0C with July as the warmest month, whereas the coolest month is January. The average annual

rainfall from 1,900 to 2,100 mm and mean annual relative humidity is 82% (Bui et al. 2016; Indochina. 2004).

c. Hydrology

Due to characteristics of topography which is well-developed karst landscapes, drainage patterns are complicated by subterranean passages and most of the streams on the island are seasonal, forming after tropical storms, except some streams situated in the higher valleys. The centre of Cat Ba island is no more than 5km from the coast, and surface drainage is poorly developed and seasonal (Indochina. 2004; Bui et al. 2016).

d. Vegetation

Cat Ba National Park supports a diversity of natural habitats with limestone forest as the major natural vegetation type. Cat Ba island had a significant biodiversity value as it is home to some rare and endangered species of plants and identified as one of the areas of highest biodiversity importance in Vietnam (Brooks 2006).To date, more than 800 vascular plant species have been described at the national park, including 25 species listed in the Red Data Book of Vietnam (1997). The forest resources have been exploited intensively by the inhabitants for timber, fuelwood, honey, bamboo shoots, and edible roots. In recent years, tourism has come central to the local economy, also pose a threat to the island's environment (Tan, Thu & Dell 2012; Bui et al. 2016).

3.3.3 Phu Tho province

a. Topography and geology

Phu Tho is a mountainous midland province is located in north-eastern Vietnam, between $104^0$48' to $105^0$27' E and $20^0$55' to $21^0$43"N, covers an area of 3,519 km^2 and about 80km far away from Ha Noi. The province is located at a joining point among three main rivers of northern Vietnam (Red River, Da River and Lo River). The area is mainly covered by low hills with the slope varying from 10^0 to 35^0. The altitude ranges between 60 t0 100m above sea level (Thai et al. 2010).

b. Climate and hydrology

Located in the tropical monsoon region, the area climate is distinguished by dry and rainy seasons. The average annual rainfall is 1,800 mm per year. The mean temperature ranges from 22.2^0C to 26.5^0C, with an average of 22.9^0C. The relative humidity ranges from 75-88%, with an average of 83% (Thai et al. 2010). The province has a fairly advantageous hydrological condition with the pass-by of three great rivers which provide a tremendous water body for irrigation and transportation.

c. Vegetation

The original vegetation type of the province was the evergreen forest composed of several typical species of the Fagaceae and Fabaceae family and plenty of bamboo, rattan, grass and shrub species. However, the site was degraded between the 1970s to 1990s due to

over-exploitation, illegal cutting, shifting cultivation and grazing and removal of the topsoil layer. Recorded growth in plantation has been very poor as species in plantation grew at half the rate of plants of the same species that had been planted in the fertile soils (Tran 2001).

3.3.4 Hoa Binh province

a. Topography and geology

Hoa Binh province is located in the north-western region of Vietnam, spreads in an area of approximately 4,660 km^2 between 104°48′E to 105°50′E and 20°17′N and 21°08′N, covering a total area of 4,698 km^2. The elevation in the province ranges from 0 to 1,510 m above sea level and gradually decreases from northwest to southeast. The topography ranges from valley to gentle slope with the angle from 20^0 to 35^0. The province's landscapes can be classified into three basic groups: the mountainous complex, the hilly complex, and the valley with a diverse topography such as mountains, small valleys, hills, cliffs and plains (Thai et al. 2010).

Geologically, the area comprises limestone, conglomerate, aphyric basalt, sandstone, silty sandstone, and black clay shale. Most parts of the area are dominated by soils such as ferrallisols, fluvisols and acrisols, which are the remnants of ancient soils on slopes after exposure to severe soil erosion (Son & Binh 2015).

b. Climate

Hoa Binh province is situated in the monsoonal region, with two separately rainy and dry seasons. The hottest month is July, and the coldest month is January with an average temperature of 26.7^{0}C and 14.9^{0}C, respectively. The maximum temperature is about 42^{0}C in the summer and minimum 3^0 C in the winter. The rainy season is generally from May to October with a high frequency of intense rainfalls which is rough 200 mm per month, with the peak in August and September from 300 to 400 mm per month (Thai, Lee & Woo 2010).

c. Hydrology

The province has four main river systems (Da river, Boi river, Buoi river, and Ma river) and Hoa Binh dam which play an important role in the irrigation system of the area and also for whole Red river delta at downstream. Due to the fragmentation of the topography, the river and streams systems usually slope and fragmented. As a result, flood usually occurs in the rainy season and causes severe damages crops and infrastructure (Hoa Binh Committee report 2010).

d. Vegetation

Forest land is the major area of the land use accounting for 52.6%, whereas barren land and non-forest rocky mountain, agricultural land, settlement areas, water surface, and grassland account for 47.4%, collectively (Thai et al. 2010). The major crops cultivated in this area include upland rice, lowland rice, maize, cassava, and sugarcane. According

to the Annual report from the Ministry of Natural Resource and Environment (MONRE) 2015, the province is featured by land use changes, soil degradation and nutrient losses associated with massive deforestation in the last 15 years, expansion of agricultural activities, and inappropriate conservation practices.

3.4 Summary and outlook

Studies on the conservation and sustainable use of both wild and cultivated plant species and their habitats have fallen far behind the declining conservation status and the demand for the important resources, particularly in developing countries. Each taxonomic group has unique ecological and socio-economic association that must be appreciated to attain an effective utilisation and conservation. In particular, secondary metabolites are restricted to particular taxonomic groups such as family, genus and even species. The species *C. hindsii* is well-known to produce various types of sesquiterpenes, alkaloids, triterpenes, diterpenes, and flavonoids (Hu et al. 2014). Due to lacking a comprehensive study of *C. hindsii*, further phytochemical analysis of this species is necessary, which will optimise the utilisation of potentially important medicinal germplasm resources especially with additional information from genetic diversity of the wild and cultivated populations (Mu et al. 2017). Genetic diversity investigation also is valuable to understand population size, the degree of isolation and fitness to save endangered plant species from the risk of extinction, particularly for those scattered in small and fragmented populations like *C. fugax*.

CHAPTER 4

GENETIC DIVERSITY

4.1 Introduction

The significant role of plant diversity in human survival and the environment has been recognised globally, not only because of increasing demand of human beings and food for everyone but also because of the diversity and fitness of flora has become more and more threatened, and many plants are becoming extinct in the wild. Effective conservation of those wild plants and germplasms to ensure the provision of nutrients, food security, and traditional medicine, as well as a healthy ecosystem, is crucial to prevent the further depletion and extinction; particularly for rare and endangered plant species. Up to date, there have been many approaches to conserving those important plants adopting different strategies (*in situ* or *ex situ* conservation) from the most problematic concerns in the tropics to temperate areas of the globe (Corle 2016). Studies have mostly involved facilitating protected areas, controlling overexploitation and *ex situ* conservation activities, which often require time-consuming and costly inventory and conservation status assessment. However, the inconsistencies of these projects have been reported, and the full genetic diversity of a species has not been guaranteed to be conserved. Recently the perception that conserving a species may not merely grow enough number of individuals but more importantly to preserve the genetic diversity of that species. Therefore to formulate appropriate management strategies oriented towards conservation, the species' biological characteristics and their environmental vulnerability must be provided, as well as knowledge of genetic variation between and within populations needs to be examined; usually now by the assistance of molecular techniques (Wee et al. 2015).

Genetic diversity represents the heritable variations distributed within and among populations of a species, migrates through generations and is encoded in the DNA of an organism, which is regulated by some evolutionary processes such as mutation, selection, genetic drift and species mating system. Along with these processes, genetic diversity is often degraded as the consequences of some biotic and abiotic conditions which directly affect the survival of wild populations. In cultivated populations, genetic diversity is greatly affected by selection, breeding system and crop conditions of plant accessions. Therefore, the role of maintenance of genetic diversity is not only to keep the wild population away from the risk of extinction for long-term survival but also to maintain a healthy cultivated plant system for selective breeding purposes.

C. fugax is one of the rarest cycad species of family Cycadaceae which is assessed as critically endangered (CR) in 2010 ICCN Red list (Osborne et al. 2010; Osborne et al. 2007) and Vietnam Data Redbook 2007 (Ban et al. 2007). The leftover representatives of this species are only found in some extremely small isolated populations in Phu Tho province, Vietnam with roughly 200 individuals distributed at an altitude of 200 m above sea level. Most of the *C. fugax* populations have been reduced (by number and size) by

the destruction of natural habitats for agricultural and plantation encroachment (Osborne et al. 2010). With declining populations and extremely narrow distribution of *C. fugax*, this locally endemic species should be assigned a high priority for conservation. Therefore, understanding of the genetic variation within and between populations is essential for the establishment of effective and efficient conservation practices for this critically endangered species.

C. hindsii (Celastraceae) is a scandent shrub widely distributed up to an altitude of 2500m through most parts, but mainly in forests, thickets and mountainous regions of China, Northern Vietnam, India, Malaysia, and Myanmar. In Vietnam, *C. hindsii* was found to be distributed in different wild populations in Phu Tho, Ha Noi, Hoa Binh and Hai Phong, and have recently been planted in many other places (according to the study investigation). Previous studies on this species revealed a variety of bioactive chemical components including sesquiterpene, triterpenes, alkaloids, and flavonoids (Yao-Haur Kuo et al. 1995; Huang et al. 2000; Hu et al. 2014).

Plant material (stem, leaves, and seeds) from *C. hindsii* have been used for generations in Northern Vietnam, particularly by Muong people for treatment of ulcers, tumours, and inflammation, and for manufacturing of tea products (Ly et al. 2006). Since the medicinal property and the traditional use were proven scientifically, the wild populations of *C. hindsii* have been overexploited, and plants are hardly found in the wild (according to our surveys). However, their cultivated forms were widely spread in home gardens and nurseries for commercial purposes because they are successful and easily regenerated by either vegetative propagation or sexual reproduction (Loan 2012; Nam, Thang & Tho 2013). Genetic diversity of *C. hindsii* may be adversely affected due to their decrease numbers in the wild and unsystematic selection for cultivation which chiefly is based on morphological traits but not genetic diversity information. As a result, the maintenance of genetic diversity is one of the major objectives for conserving genetic variation among wild populations and beneficial in developing agricultural practices to circumvent genetic erosion and propose conservation strategies for the longstanding vigour of the species (Ismail et al. 2016).

The genetic diversity assessment at a population level and species level can be performed through a number of techniques such as morphological, biochemical and molecular analysis. However, molecular techniques have been preferred recently because of their speed, low cost and require a small amount of plant material; which make the methods desirable (Govindaraj et al. 2015). Out of the molecular techniques available, RAPD-PCR and RAMP-PCR have necessary features and have been used widely to determine the genetic relationships within and between populations for various endangered plants species (Fu et al. 2013; Rasouli & Karimi 2015; Long et al. 2015; Ismail et al. 2016) and medicinal plant species (Al-Rawashdeh 2011; Kazemi & Hajizadeh 2012; Tahira & Saleem 2013; Bishoyi et al. 2016; Baruah et al. 2017). In these techniques, short oligonucleotides of arbitrary sequence (in RAPD-PCR), and combination of RAPD and anchored microsatellites primers (in RAMP-PCR) are used to support the amplification

of regions of selected plant genomes, and these PCR products are separated by size in gel electrophoresis. The variation between genotypes is reflected as differences in the banding patterns.

Currently, there are no reports on the genetic variation of *C. fugax* and *C. hindsii* using any markers, although some studies have used molecular markers on other species of Cycad (Xiao et al. 2004; Xiao & Gong 2006) and a related *Celastrus* species (Prasad 2007). Therefore, the objectives of this study are to provide information on the genetic diversity of these two species, using two different molecular markers for variations within and between populations. The study, therefore, contributes valuable information for conservation and of the two species, in particularly, a rescue plan for the critically endangered *C. fugax*.

4.2 Material and methods

4.2.1 Study site

C. fugax has strictly small and fragmented populations either in the wild or cultivated condition with a few individuals left in one commune of Tram Than, Phu Tho Province in the north of Vietnam (Osborne et al. 2010). The original populations of this species have been almost entirely cleared over time by acacia and eucalyptus plantations. However, several individuals have been collected from the wild and planted in home gardens as specimens or decoration purposes (Ban et al. 2007).

Samples of *C. hindsii* were collected from various locations in four provinces across the north of Vietnam (Figure 3.4). Of those areas, samples only from Hai Phong City (in Cat Ba National Park population) were collected from their original habitats whereas the rest were collected in cultivated conditions such as home gardens, intensive farming (Phu Tho, Hoa Binh populations), and buffer zone of Ba Vi National Park (Ha Noi population).

To test the hypothesis that genetic diversity varies between populations of the two species, the study sites were classified into different sampling areas divided into natural forest or cultivation, mountain or lowlands and restricted areas or buffer zone (if the samples were collected from a national park).

4.2.2 Sampling strategy

Plants to be used in this project were collected and ecological parameters assessed on site in Vietnam; including population size and fragmentation. Healthy growing individual plants within larger populations were selected randomly (for *C. hindsii*), however, in the area with a small number of individuals, most of the plants were collected (*for C. fugax*). To ensure statistical reliability of the result, representative samples of the genetic variation were collected and named by using sufficient numbers where possible. Materials were assessed and permitted by the Australian Quarantine and Inspection Service (AQIS) as covered by the criteria of Condition C17727, Schedules 4 and 6 of the Quarantine

Proclamation 1998 and stored in an Approved location for Quarantine stock in a cold room at the laboratory of the University of Technology, Sydney (UTS). Locations and description of the sample sites of the two species are presented in Table 3.1 – 3.2.

The distribution of *C. fugax* has been dramatically decreasing, and according to the estimation of Osborne et al. (2010), only 200 *C. fugax* occurred both in the wild and planted in home gardens near its original habitats. However, in our case, a total of only 30 *C. fugax* individuals were found and collected in Tram Than commune, Phu Ninh district, Phu Tho Province, Vietnam. Of those, leaves from 16 individuals representing four different populations, which were classified according to their location characteristics (Table 3.1), were selected for the study. Of these four populations, three are cultivations in home gardens named Lam village 1 (LV1), Lam village 2 (LV2), and Than village (TV); and one grows naturally in the forest (NF).

C. hindsii has been overexploited in the wild. However, they are widely cultivated in home gardens, nurseries and intensive farming for research and commercial purposes all over the provinces in the north of Vietnam. The reputation of the species has been highlighted in Hoa Binh province where folk medicine practices and commercial cultivations of this species have brought the residents significant economic benefits and an enviable reputation. Normally more leaf tissue was obtained than needed; however, the older counted leaves were omitted due to older plant tissue being generally unsuitable for DNA extraction. A total 16 individuals were used in the study, of which 12 were from cultivations in three provinces of Hoa Binh (HB), Phu Tho (PT), and Ha Noi (HN); and four were from natural forest in Cat Ba National Park in Hai Phong city (HP). Two out of 16 plants were classified as narrow leaf phenotype (both collected in PT) versus broad leaf phenotype of the rest of the material. To the author's best knowledge, the population found in the natural forest and studied in this research was the first time mentioned in the literature. The description of these four locations is presented in Table 3.2.

4.2.3 Sample collection and storage

There is some general agreement on how to select appropriate parts of a plant to obtain pure DNA with high quantity. According to previous studies, young and expanding leaves and flower buds yielded the best quality and quantity of DNA. Also, unexpanded leaves contain less contaminant such as polysaccharides and polyphenols, especially in case of medicinal plants. Therefore, the study primarily collected young and expanding leaves of both species *C. hindsii* and *C. fugax*. However, in the case of *C. fugax* mature leaves also were collected as no other suitable juvenile material was available on several plants.

Leaf samples were removed from each plant using a scalpel sterilised in weak bleach not only to prevent cross contamination if any of the individuals sampled carried pests or disease but also to ensure adherence to quarantine procedures. To reduce the possibility of contamination of the plant DNA with DNA from other organisms, only healthy leaf samples with no evidence of deformity or pest infestation were collected. Each sample was inspected visually for the presence of external contaminants (such as insects or micro-

organism), then wrapped in aluminium foil and placed in vacuum laminated clear bag containing silica gels for transport back to the laboratory before being stored at -80^{0}C and DNA extraction. Silica gel was identified as an ideal material for field preservation of leaf samples for DNA studies (Chase & Hills 1991).

4.2.4 DNA extraction

A DNA extraction procedure based on CTAB using the modified version of the method described by (Doyle & Doyle 1987a) has proven to be successful over a number of species of plants and is well proven to maximise DNA amounts from woody samples (Magel et al. 2002). CTBA is a positively-charged detergent that allows the extraction of DNA from plant material with a minimum of containing polysaccharides. The DNA was in most cases suitable to be used for RAMP-PCR and RAPD-PCR (De Filippis & Magel 1998).

Before the DNA extraction process, all solutions (except heat labile or flammable compounds) and equipment used were sterilised by autoclaving (121^{0}C, 104kPa for 20 min) or by washing (the mortar and pestle only) in 10% sodium dodecyl sulfate (SDS). Fresh CTAB buffer including 2% (w/v) CTAB, 20mM EDTA, 100mM Tris-HCl, pH 8.0, 1.4 M NaCl, 1% (v/v) 2-mercaptoethanol (2-ME) and 1% (w/v) polyvinylpyrrolidone PV-40 (PVP) and was prepared on the day of each extraction. The sequential DNA extraction protocol for both study species was implemented in the following steps:

(i) Preparation of tissue lysate

1. Pre-cool a mortar and pestle in liquid nitrogen for several mins.
2. Grind 100mg of leaves to a fine powder in liquid nitrogen.
3. Add 450µl CTAB buffer to the grinding process and make the mixture thaw then add a further 450µl CTAB buffer.
4. Transfer the whole mixture to a 2ml microcentrifuge tube using a wide-bore pipette and heated at 60^{0}C for 5 min in a dry heat block.

(ii) DNA purification

5. Add 6 µl of 20mg/ml proteinase K (Promega) to the mixture and further incubate the tube in a water bath at 37^{0}C for 90 min with occasional inversion.
6. Add 900 µl chloroform: isoamyl alcohol (24:1) to the slurry (to remove contaminants), and gently invert 8-10 times to maximise DNA yield.
7. Centrifuge the tube at 12,000 x g for 5 min at room temperature.
8. Transfer carefully the aqueous supernatant (roughly 500 – 700 µl, containing the DNA) to a fresh 1.5 ml micro-tube using a wide-bore pipette.

(iii) DNA precipitation

9. Add 2/3 the volume of cold isopropanol to the supernatant to precipitate the DNA.

10. Mix the tube gently, then centrifuge at 12,000 x g for 1 minute at room temperature to pellet the precipitate.

11. Pour gradually 1 ml chilled 70% (v/v) ethanol, and the DNA pellet should be gently disrupted with the side of a pipette tip to maximise the surface area available for washing

(iv) DNA washing

12. Centrifuge the DNA pellet at 12,000 x g for 10 min at room temperature.

13. Decant ethanol, and then dry the DNA pellet at room temperature for a maximum of 15 min.

(v) DNA re-solubilisation

14. Add 50 µl sterile TE buffer (10mM Tris-HCl, 1 mM EDTA, pH 8.0) to suspend the DNA.

15. Incubate the contents in a water bath at 37^0C then left overnight at 4^0C (for fully dissolved of the DNA pellet).

(vi) RNA removal

16. Add 2 µl of 10 mg/ml RNAase A (Sigma) and incubate in a water bath at 37^0C for 30 min to destroy any RNA contaminant. Finally, the DNA solution should be analysed by spectrophotometers and stored at -20^0C for later use.

4.2.5 DNA qualification

The quantity and quality of extracted DNA was estimated using a Pharmacia GeneQuant DNA/RNA calculator. Absorbance readings were taken at wavelengths of 230, 260 and 280 nm. The purity of the DNA was then estimated by the ratio of readings taken at 260:280nm and 260:230nm (pure DNA having a 260:280 ratio of 1.8 - 2.0). The calculator automatically estimated DNA yield by multiplying the absorption at 260nm by 50 µg/ml (an OD_{260} of 1 corresponds to ~ 50µg/ml of double-stranded DNA).

4.2.6 Polymerase Chain Reaction

DNA polymorphisms in the genome of the two selected plant species (*C. hindsii* and *C. fugax*) were assessed using the Polymerase Chain Reaction (PCR), and a variety of RAPD primers and microsatellite primers. Each PCR was performed in a final volume of 25µl containing 1 x Fisher Biotech Reaction Buffer (67mM Tris-HCl - pH 8.8, 16.6mM $[NH_4]_2SO_4$, 0.45% Triton X-100, 0.2mg/ml gelatin), 1M betaine, 2mM Fisher Biotech mixed dNTPs, 0.1µg/ml RNase A, and varying amounts of primer/s, $MgCl_2$, Fisher Biotech Taq DNA Polymerase, genomic DNA, and sterile double distilled water. PCR reactions were carried out using a Biometra Personal Cycler.

4.2.7 PCR preparation

To prevent contamination of the PCR reaction mixture with exogenous DNA, and to prevent cross-contamination of template DNA, the following precautions were taken as recommended by Hu (2016);

1. All equipment used in PCR preparation (PCR tubes, pipette tips, master mix tubes, and solutions) were autoclaved by appropriate cycles in Systec DX-150 (Germany).
2. Laminar flow cabinet, pipettes, and gloves were sterilised by irradiation with UV light before sample preparation;
3. Reaction components were assembled in the laminar flow cabinet;
4. Gloves were worn, and were changed periodically during PCR preparation;
5. Reagents were pre-mixed, before dividing into aliquots, with template DNA added last, and
6. A negative control (containing all components except template DNA) was included in each series of reactions to ensure no contamination DNA was present in any of the reagents.

Test reactions were performed in duplicate to ensure the reproducibility of the results.

4.2.8 Primer selection and optimisation of PCR

Primer selection and optimisation of the PCR are described in Appendix 1. Final reaction conditions:

PCR was carried out on each of the 16 samples for each species either *C. hindsii* or *C. fugax* with RAPD primers. DNA samples were also assessed using different combinations of one microsatellite primers (either MS1: 5' CAACACACACACAC 3' or MS2: 5' TGACACACACACAC 3'; or MS3: 5' CAACTCTCTCTCTCT3') with one of the various RAPD primers. Reactions were carried out in a Biometra Personal Cycle using the following temperature profile: Denaturing at 94^0C for 1 min, then 46 cycles of denaturation at 92^0C for 1 min, annealing 35^0C (39^0C or 41^0C for RAMP-PCR) for 1 min, extension 72^0C for 1 min, finally an extension at 72^0C for 5 min and cooling to 4^0C.

4.2.9 Gel electrophoresis

The above PCR products were analysed by electrophoresis on agarose gels.

a. Gel preparation

Agarose gels of 2% (w/v) were prepared using Sigma Type I (low EEO) agarose dissolved in 0.5 x TBE buffer (5.4 g/L Tris base, 2.75 g/L boric acid, 2 ml/L (0.5 mol/L) EDTA pH 8) by heating in microwave stove for 30 seconds twice. Each gel was cast and run in a Biorad Mini-Sub® Cell electrophoresis tank. The gels were loaded with 15µl of each PCR product combined with 2µl of bromophenol blue loading dye (0.2% (w/v) bromophenol blue and 50% (v/v) glycerol in TE buffer). A molecular weight marker

(Hyper Ladder I – Biotech) was added to one lane to determine the approximate sizes of the PCR products. Electrophoresis was carried out at 90V, in 0.5 X TBE buffer, for two to three h till the dye had reached close to the end of the gel.

b. Gel staining and photography

Agarose gels were stained by GelRed™ by diluting the GelRed™ 10,000X stock reagent into the agarose TE buffer solution at 1: 10,000 (e.g, 5 µl of the GelRed™ 10,000X stock reagent added to 50ml of the gel solution). The GelRed solution was mixed thoroughly by swirling. After complete running gels, there were stained with GelRed dye and viewed and photographed by E-gel Imager connected to a computer. Images were printed in black and white using a Cannon-iR4251 laser printer.

4.2.10 Band scoring

The presence and absence of bands were determined by examining each gel photograph. Observed bands were marked in ink on one of two photographs taken of each gel to obtain a permanent record for later evaluation. Only those bands that were unambiguous were scored. The observed amplicons were scored '1' for the presence and '0' for the absence of bands, and the binary data used for statistical analysis. The band sizes were determined by comparison with a 100 bp DNA molecular weight ladder (Promega), and faint bands of doubtful reproducibility were ignored.

4.2.11 Statistic analysis

The binary data was analyzed with PopGen Version 1.31 (Yeh et al. 1999), a Microsoft Windows-based freeware program for population genetic analysis and PRIMER Version 5 (Clarke & Gorley 2001) to determine genetic parameters such as Nei's genetic diversity (H), Shannon's diversity index (I), diversity within population (Hs), total gene diversity among populations (Ht), gene flow (Nm), coefficient of gene differentiation (G_{ST}), and genetic distance Jost (D). The G_{ST} estimates genetic differentiation and reduction in the number of heterozygote loci based on Nei's regular and unbiased genetic measures (Yeh et al. 1999). The Mantel test statistic (r) was used to determine the correlation between geographic and genetic distances using the program IBD Isolation by Distance Version 1.52 (Mantel 1967). Mantel (r) test typically compares two distance matrices that were calculated for the same set of objects but are based on two independent sets of variables (e.g. a species dissimilarity matrix and population site distance matrix). The test calculates the correlation between values in the corresponding positions for two matrices. The significance of the linear relationship between matrices is assessed through basic permutation statistics.

Principal components analysis (PCA) can be used as a simple visualisation tool to summarise dataset variance and show the dominant gradients in low-dimensional space. PCA results are usually displayed as a two- or three-dimensional scatter plot, where each axis corresponds to a chosen principal component, and each object is plotted based on its

corresponding PC values. Multi-dimensional scaling (MDS) was used to understand patterns of variation within and amongst populations by converting a set of variables into a few dimensions so that individual variations are condensed into a set of limited axes. MDS is a unique ordination technique in that a small number of ordination axes are explicitly chosen before the analysis, and the data are then fitted to those dimensions. Thus, if only 2 or 3 axes are chosen, there will be no nondisplayed axes of variation at the end of the analysis. Similar to PCA, a matrix of object dissimilarities is first calculated using a chosen distance metric. The goal of cluster analysis is to separate variables into groups based on the similarity of the variable scores among objects, so that variables within each group (cluster) are more similar to one another than to variables in other groups. The algorithms used usually minimise the within-group distances and maximise between-group distances. Such graphical analysis helped to identify the individuals and primers which tend to cluster together. Cluster procedure was an average linking one, and all similarities used were Bray-Curtis to produce dendrograms (Clarke & Gorley 2001).

ANOSIM and ANOVA are statistical tests of significance ANOSIM tests for the significant difference between two or more classes of objects based on any (dis)similarity measure (Clarke & Gorley 2001). It compares the ranks of distances between objects of different classes with ranks of object distances within classes. The basis of this approach is similar to the MDS ordination technique described above. As ANOSIM is based on ranks, it has fewer assumptions compared to regression techniques such as analysis of variance (ANOVA).

The power of meta-analysis relies on very specific molecular methodological and statistical treatment and bioinformatics of the individual studies. In a practical meta-analysis, F_{ST} and analogues such as genetic differentiation (G_{ST}), Nei genetic diversity (H) and Jost differentiation statistic (D) (Nei 1973; Jost 2008, 2009) are somewhat constrained by expected within-population heterozygosity (Meirmans & Hedrick 2011), which may hinder cross-study comparisons. Despite this we recommend the use of standardised analogues G_{ST}, H and D, as meta-analysis may be affected by size. However, these indices are the best way to alleviate the confounding effects of different maximum possible values across studies (Jost 2008, 2009; Leng & Zhang 2011). Finally, we conclude that $G_{ST,}$ H and D meet the criteria and consistency for including effective size differences and avoid the difficulty of calculating errors for these values (ArchMiller et al. 2015)**.**

4.3 Results

4.3.1 DNA extraction

Total genomic DNA mini-preparations were conducted employing a modified protocol of the CTAB-DNA extraction method of Doyle & Doyle (1987). It is believed that better quality and a larger quantity of DNA are obtained from extraction using young fresh meristematic leaf material homogenised in liquid nitrogen (Tiwari et al. 2012). This was also confirmed by this study when comparing quality and quantity of DNA extracted from

young leaves and mature leaves of *C. hindsii* - both in dry condition. The results from Table 4.1 show that young leaf provided not only a higher yield of DNA, but also better quality as indicated by the $A_{260:280}$ ratios and a more constant $A_{260:230}$ ratios around 1.6 and 1.8 in young leaves, and 1.1 and 1.12 in mature leaves, respectively. There were statistically significant differences between the age of leaf tissues and DNA yield, and DNA purity (P=0.05) according to ANOVA analysis. However, the quality of the DNA from the mature and young leaves of *C. hindsii* was not as variable as the quantity.

Table 4.1 Quality and quantity of DNA extracted from young and mature leaves of *C. hindsii*. Yields and ratios were estimated using a Genequant DNA/RNA calculator. The 260:230 and 260:280 ratios refer to absorbance measured by spectrophotometry at 230 nm, 260 nm, and 280 nm respectively

Type of leaf	260:230 ratio			260:280 ratio			DNA yield (µg/g A.D. tissue)		
	Min.	Max.	Mean ± SE	Min.	Max.	Mean ± SE	Min.	Max.	Mean ± SE
Young	1.162	2.870	1.782 0.057	1.400	1.810	1.605 0.016	530	4915	2191.094 218.789
Mature	0.424	3.000	1.121 0.161	0.840	1.655	1.098 0.046	85	3410	743.500 187.040

Although leaf materials from *C. fugax* are hard fibrous and mature, its DNA quality was not significantly lower as compared to the young leaves of *C. hindsii* using the same protocol with a modification of CTAB method (Doyle & Doyle 1987). With the average $A_{260:280}$ ratios of about 1.6 and 1.8, respectively, the DNA quality of *C. fugax* is was better than that of *C. hindsii.* However, there was a small difference in DNA yield between the two species (Table 4.2).

Table 4.2 Quality and quantity of DNA extracted from mature leaves of *C. fugax* and young leaves of *C. hindsii*. Yields and ratios were estimated using a Genequant DNA/RNA calculator. The $A_{260:230}$ and $A_{260:280}$ ratios refer to absorbance measured by spectrophotometry at 230 nm, 260 nm and 280 nm respectively

Species	260:230 ratio			260:280 ratio			DNA yield (µg/g A.D. tissue)		
	Min.	Max.	Mean ± SE	Min.	Max.	Mean ± SE	Min.	Max.	Mean ± SE
C. fugax	1.636	2.74	2.064 0.046	1.651	2.381	1.841 0.027	685	4195	2045.741 141.665
C. hindsii	1.162	2.870	1.782 0.057	1.400	1.810	1.605 0.016	530	4915	2191.094 218.789

Leaves of *C. hindsii* are believed to contain excessive amounts of secondary metabolites and phenolic compounds which promptly turn the leaves a brown colour when they are exposed to oxygen and high temperature. The DNA extracted from this browning material

was contaminated and low in DNA. The $A_{260:280}$ ratio was on average around 1.6 with the lowest sample at 1.4. However, these contaminants did not affect PCR and gel electrophoresis as reproducible profiles have been acquired from the DNA in our study. It is also in agreement with De Filippis & Magel (1998) and Hung et al. (2011); as proteins contaminated DNA samples with a low $A_{260:280}$ ratios (about 1.5).

4.3.2 PCR optimisation

Optimization is a very important step to obtain a high diversity of DNA bands (Omalsaad et al. 2014). The study optimised the RAPD-PCR and RAMP-PCR to obtain consistent, reproducible and sharp amplicons by adjusting PCR reaction parameters such as amounts of DNA, *Taq* polymerase, $MgCl_2$, and annealing temperature (as depicted in Table 4.3). The result was on an agreement with genetic diversity study of Hung, De Filippis & Buckney (2011) on several endangered forest trees using RAPD and RAPM-PCR.

Table 4.3 Optimization of PCR parameters

PCR parameter	Tested range	Optimum condition	
		RAPD-PCR	RAMP-PCR
DNA concentration	10, 15, 20, 25, and 40ng	25ng	25ng
***Taq* polymerase**	0.5, 1.0, and 1.5 U	1 U	1.5 U
$MgCl_2$ (25	1.5, 2.0, 2.5, and 3.0 mM	2.5 mM	3.0 mM
Annealing temperature	35, 37, 39 and 41^0C	35^0C	39 or 41^0C

The concentrations of *Taq* polymerase, $MgCl_2$, and annealing temperatures were optimised using template DNA (10, 15, 20, 25, and 40ng) to begin the optimisation process to find the best of DNA banding pattern. It appeared that 25ng of DNA provided the most diversity and the sharpness of band. Optimization of *Taq* polymerase was conducted with optimal template DNA amount using 0.5, 1.0, and 1.5 units, and the results showed that 1.0 unit and 1.5 units of *Taq* polymerase was optimal for the tested RAPD primers and RAMP primers respectively. However, there were no significant differences in banding pattern of DNA in gels using 1.0 and 1.5 units of *Taq* polymerase gained from RAPD-PCR, the amount of 1.0 unit was selected because of economic purpose. In case of RAMP-PCR, due to a higher amount of primer combination, a greater amount of *Taq* polymerase (1.5 units) showed the best results. The greater amounts of *Taq* polymerase were not further tested because acceptable results were obtained, and more than 1.5 units have rarely been used in RAPD and RAMP-PCR.

Similarly, the optimal amount of $MgCl_2$ was depended on the particular primer conditions. In the current study, 2.5 mM and 3.0 mM of $MgCl_2$ of were found to show the most consistent banding pattern in RAPD and RAMP-PCR respectively in both plant species of *C. fugax* and *C. hindsii.* Lower than 2.5 mM and higher than 4mM of $MgCl_2$ generated smearing and faint bands which are not reproducible. Also, 37^0C was the best annealing temperature in RAPD-PCR, and RAMP primers produced the most consistent

banding pattern at 39^0C and 41^0C. The lower annealing temperature of RAPD primers in this study was consistent with most of the studies. However, there are reports where temperature as high as 41^0C (Te-chato 2000), and much higher at 50^0C and 55^0C (Sahasrabudhe & Deodhar 2010; Fernandez et al. 2003).

The optimal conditions as depicted in Table 4.3 were found to produce reproducible bands not only for fingerprints in optimised reaction experiments but also for the most parts on the two species. However, due to analytical variations such as the use of DNA samples through freeze/thaw cycles, insignificant differences were recorded and remedied by preparing fresh aliquots before each experiment.

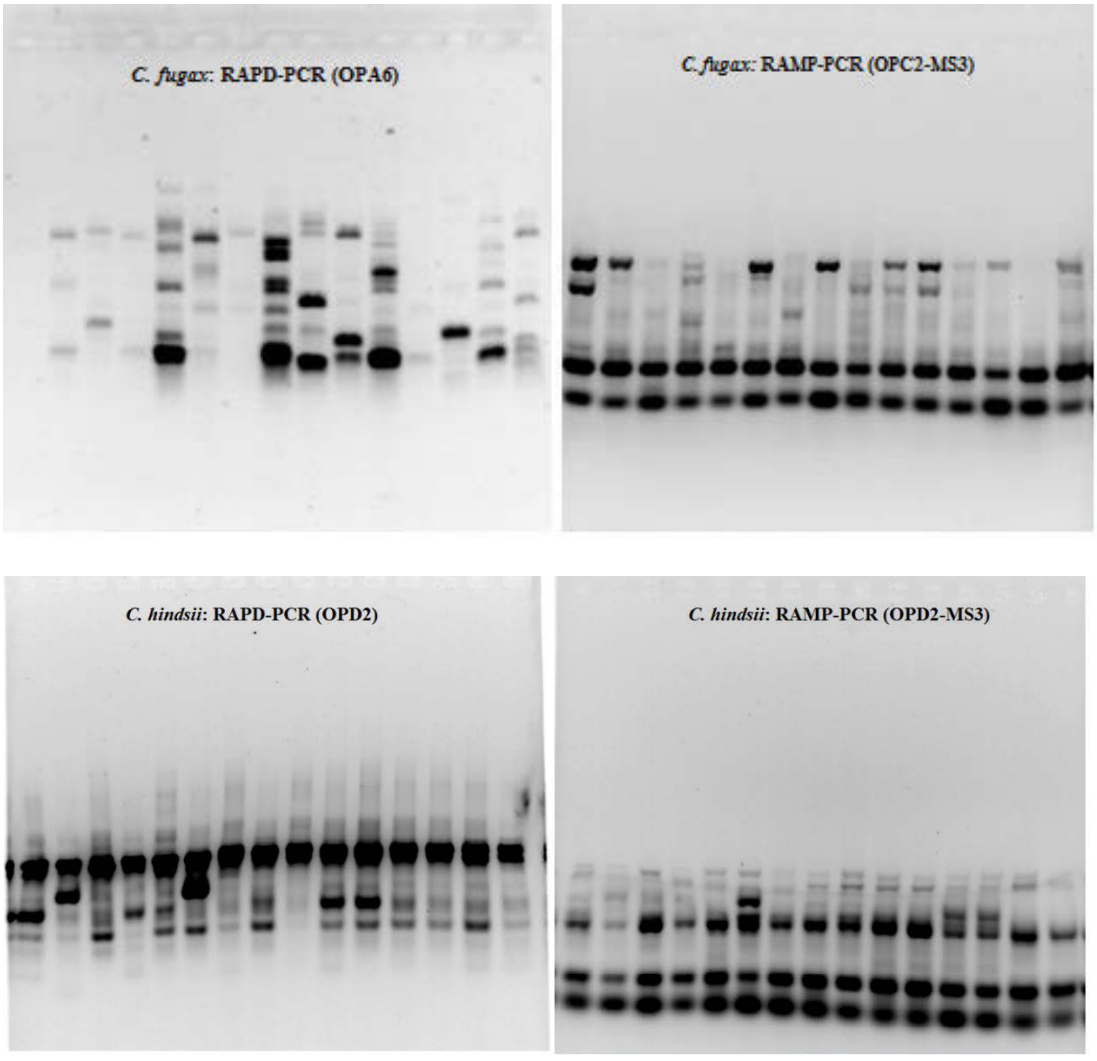


Figure 4.1 Example of a RAPD-PCR and RAMP=PCR profile among *C. fugax* and *C. hindsii* individuals

The genetic diversity within *C. fugax* and *C. hindsii* was preliminarily screened by 20 RAPD primers. Out of 20 primers used for initial screening, two sets of six primers for each species were selected for further analysis (Appendix 1), and the rest were deselected

as they did not produce any amplification. The use of microsatellites MS1, MS2, and MS3, in combination with some RAPD primers, produced more complex banding patterns (Figure 4.1). The use of two different markers (RAPD and RAMP) in this study was used to compare the effectiveness of each genetic marker in the DNA fingerprinting of two species. All tested primers for two species produced reproducible bands.

Overall results from *C. fugax* showed that six RAPD primers amplified a total of 62 bands, out of 46 were polymorphic for the entire data set. An average of 10.33 bands per primer was generated. The total number of polymorphic bands produced per primer ranged from 1 (OPC-2 and OPD-2) to 19 (OPA-6), while only two primers (OPA-6 and OPB-13) resulted in 100% polymorphism (Appendix 1). Likewise, six RAMP primers produced a total of 66 bands (an average of 11 bands per primer), of which 48 were polymorphic. While OPD-2/MS1 generated only two polymorphic bands (25% polymorphism), OPC-2/MS2 yielded the maximum of 14 polymorphic bands (93.33% polymorphism).

Unlike *C. fugax*, RAPD-PCR and RAMP-PCR used DNA materials from *C. hindsii* yielded a much higher number of total bands and polymorphic bands (Appendix 1). Six RAPD-primers amplified a total of 120 bands, out of 98 were polymorphic with an average of 20 bands generated by per primer. The total number of polymorphic bands produced per primer ranged from 8 (OPD-2) to 30 (OPD-7) equivalent to 72.73 to 93.75%, respectively. Unlikely, five RAMP primers produced a total of 69 bands (average 13.8 bands per primer), of which 47 were polymorphic. RAMP primers created a relatively similar pattern of polymorphic bands compared to RAPD primers as these values ranged from 8 to 11 per primer, equivalent to 57.14% to 82.81% polymorphism. Overall percentage polymorphism shown by RAPD and RAMP primers was 74.19% and 72.73% for *C. fugax*, and 81.67% and 68.12% for *C. hindsii*, respectively (Appendix 1).

Polymorphism among the populations of *C. hindsii* was different as indicated by RAPD primers and RAMP primers. RAPD primers showed the higher percentage polymorphism (81.67%), while RAMP primers showed only 66.67%. In contrast, both RAPD and RAMP primers detected almost the same number of polymorphic bands in *C. fugax.* Interestingly, the combinations of microsatellites (MS1, MS2 and MS3) with RAPD primers detected more polymorphic bands than that by respective RAPD primers alone in *C. fugax*, whereas in *C. hindsii* the situation was opposed as RAPD primers alone produced more bands than their combinations with microsatellite primers.

4.3.3 Intra-population genetic diversity

a. C. fugax: RAPD - PCR

16 individuals of *C. fugax* were sampled in the field study in 4 populations (Table 3.1). The actual number of plants collected was higher than the tested samples. However, several samples were omitted from the analysis as they do not represent the populations or are geographically close to the already selected individuals. On the other hand, in NF

and LV1 population, only four individuals were found at each site, all of them were selected for genetic analysis.

C. fugax was tested with six RAPD primers (OPA-6, OPB-13, OPC-2, OPC-13, OPC-19, and OPD-2). The total of bands scored was 61, of which OPA-6 generated 19, ten by OPB-13, seven by OPC-2, ten by OPC-13, nine by OPC-19 and seven by OPD-2. Overall all genetic parameters such as the percentage of polymorphism, Shannon index (I) and Nei's gene diversity (H) were elevated in LV2 and NF with the highest at 46.77%, 0.24 and 0.17, and 45.16%, 0.24 and 0.16, respectively. While TV had the lowest results of polymorphic bands (22), and this was also reflected in % polymorphism of 35.48%, Shannon index of 0.19 and Nei's gene diversity of 0.13 (Table 4.4). TV population was also isolated within a small hillside habitat in which inbreeding might have occurred, and this ecological condition may lead the low genetic diversity in TV.

Table 4.4 DNA polymorphism between four populations of *C. fugax* detected by RAPD-PCR. Six RAPD primers were used. H is Nei's (1973) gene diversity. I is Shannon's Information index (Lewontin 1972)

Parameter	LV1	LV2	NF	TV
Number of individuals	4	4	4	4
H	0.14	0.17	0.16	0.13
I	0.21	0.24	0.24	0.19
Number of polymorphic bands	24	29	28	22
Percentage of polymorphism (%)	38.71	46.77	45.16	35.48

b. C. fugax: RAMP-PCR

NF population had the highest number of polymorphic bands (29) generated by combinations of RAPD primers OPC-2 and OPD2 with microsatellite primers MS1, MS2 and MS3 as well as the percentage of polymorphism (43.94%), whereas TV population had the lowest in all aspect of analysed indices. This amplification was the same as the results produced by RAPD primers. Interestingly, the combination of microsatellite primers produced more bands per fingerprint for this species than the respective RAPD primer alone. This result agreed with Sommerville (2001) and Hung, De Filippis & Buckney (2011)'s studies in the endangered species *Grevillea caleyi* and *Markhimia stipulate*, respectively. In addition to the agreement, microsatellite primers of MS3 and MS2 generated more bands than MS1 when both combined with either OPC-2 or OPD-2.

Overall all genetic parameters such as P(%), (I) (H) were slightly different as predicted by RAPD primers, which is higher in NF and LV1 (was LV2 by RAPD primers) with the highest at 43.94%, 0.2 and 0.16, and 39.39%, 0.23 and 0.15, respectively. Similarly, TV retained the lowest results of polymorphism of 26.15 %, I of 0.15 and H of 0.09 (Table 4.5).

Table 4.5 DNA polymorphism between four populations of *C. fugax* detected by RAMP-PCR. Five combinations of two RAPD primers (OPC-2, OPD-2) and three microsatellite primers (MS1, MS2, MS3) were used. (H) is Nei's (1973) gene diversity. (I) is Shannon's Information index (Lewontin 1972)

Parameter	LV1	LV2	NF	TV
Number of individuals	4	4	4	4
H	0.15	0.12	0.16	0.09
I	0.23	0.19	0.24	0.15
Number of polymorphic bands	26	23	29	18
Percentage of polymorphism (%)	39.39	34.85	43.94	26.15

c. C. hindsii: RAPD-PCR

Fifteen individuals of *C. hindsii* were sampled in four populations, and an average of four individuals per population, except HP which consisted of only three individuals. The criteria for selecting populations and individuals were based on the geographic distance between individuals and populations to estimate the genetic flow between populations. Moreover, vegetation situations and terrains condition were also included in to ensure that differentiation may lead to indications of genetic diversity of each population. Therefore, all individuals collected from the population in Hai Phong city (HP) were from natural forests of Cat Ba Island which is the first time mentioned in both local and international literature. The other two populations are close to their originally distributed limestone mountain habitats with Hoa Binh population (HB) collected in farms near Cuc Phuong National park, and Ha Noi population (HN) collected in the buffer zone of Ba Vi National Park and home garden. *C. hindsii* population in Phu Tho province (PT) was collected from home gardens which all specimens were previously unrooted and migrated from other unknown places.

Table 4.6 DNA polymorphism between four populations of *C. hindsii* detected by RAPD-PCR. Six RAPD primers were used. (H) is Nei's (1973) gene diversity. (I) is Shannon's Information index (Lewontin 1972)

Parameter	PT	HN	HB	HP
Number of individuals	4	4	4	4
H	0.20	0.14	0.15	0.18
I	0.29	0.22	0.22	0.27
Number of polymorphic bands	63	46	45	54
Percentage of polymorphism	52.50	38.33	37.50	45.00

C. hindsii was investigated with six RAPD primers (OPA-2, OPA-7, OPC-2, OPD-2, OPD-7, and OPD-10). All primers tested in this study produces various patterns of bands within populations with more bands produced several primers or populations, while fewer bands generated by the others. The data from Table 4.6 shows that high polymorphisms

(about 53%) were detected within the population PT whereas much lower polymorphic bands as well as (H) and (I) were calculated in population HB. The isolated geographical condition of Cat Ba Island probably did not cause low gene flow and diversity of *C. hindsii* population in HP. This result not only reflected the furthest distance from this population to the rest but also may indicate that the primers used in this experiment are not necessarily useful primers to differentiate among between populations.

d. C. hindsii: RAMP-PCR

While OPD-2 and its combination with MS3 produced a roughly equal number of bands per fingerprint for this species (11 and 12 bands, respectively), the other three RAMP primer combinations produced fewer bands than the respective RAPD primers alone (32 bands over 16 bands between OPA-7 and OPA-7 with either MS1 or MS2; 17 bands over 11 band between OPC-2 and OPC-2 with MS3). It probably due to the RAMP primers, which are not useful to detect the polymorphism difference between populations of *C. hindsii*. Therefore, the number of polymorphic bands of all population was much lower than that produced by respective RAPD primers, from the lowest of 19 bands (in HP) to the highest of 30 bands (in PT). Furthermore, there was the difference between the number of bands produced by MS1 and MS2 as they combined with RAPD primer OPA-7.

From the data produced by RAMP primers as can be seen in Table 4.7 that PT was dominant from the rest with the highest percentage of polymorphism (43.48%) as well as H and I indexes (0.15 and 0.22, respectively), and HP remained the second highest diversity indices as produced by RAPD primers. On the other hand, HN and HB showed almost equal results of all genetic diversity indices.

Table 4.7 DNA polymorphism between four populations of *C. hindsii* detected by RAMP-PCR. Four combinations of two RAPD primers (OPA-7, OPC-2 and OPD-2) and three microsatellite primers (MS1, MS2, MS3) were used. (H) is Nei's (1973) gene diversity. (I) is Shannon's Information index (Lewontin 1972)

Parameter	PT	HN	HB	HP
Number of individuals	4	4	4	4
H	0.15	0.11	0.11	0.12
I	0.22	0.16	0.17	0.18
Number of polymorphic bands	30	20	20	21
Percentage of polymorphism	43.48	28.99	28.99	30.54

4.3.4 Inter – population genetic diversity

a. Genetic diversity

Genetic diversity parameter such as percentage polymorphism (P%), Nei's gene diversity (H), Shannon's diversity index (I), diversity within a population (Hs), and total gene diversity among populations (Ht) was shown in Table 4.8. These five parameters (P%, H,

I, Hs, Ht) predicted by RAPD data were found to be higher (74.19%, 0.24, 0.36, 0.27, 0.15 for *C. fugax* and 81.67%, 0.27, 0.41, 0.27, 0.18 for *C. hindsii*) than RAMP data (72.73%, 0.22, 0.34, 0.22, 0.12 for *C. fugax* and 66.67%, 0.19, 0.29, 0.19, 0.12 for *C. hindsii*), respectively, at species level. The P%, H, and I indices were found lower at the population level, and represent low genetic diversity at the species level but high population differentiation. Among four populations of each species, NF of *C. fugax* exhibited the highest average level (between RAPD and RAMP data) of variability (44.55% polymorphisms; I = 0.24; H = 0.16) while population TV exhibited the lowest level (30.81% polymorphisms; I = 0.17; H = 0.11). Similarly, population PT of *C. hindsii* showed the highest average level of genetic diversity indices (47.99% polymorphisms; I = 0.26; H = 0.18), while the lowest was recorded in HB (33.25% polymorphisms; I = 0.18; H = 0.13).

Table 4.8 DNA polymorphism and population statistics and Mantel test. G-statistic is identical to the F-statistic where only two possible alleles are scored at any locus, or their analysis of presence or absence of bands is conducted. G_{ST}, however, does not require knowledge of genotype frequencies and is a measure of the reduction in the numbers of heterozygote loci and genetic differentiation. Nm = estimate of gene flow from G_{ST}, Nm = 0.5(1- G_{ST})/G_{ST}. Mantel statistic (r) was used to determine the correlation between geographic and genetic distance.

Parameter	*C. fugax*		*C. hindsii*	
	RAPD	**RAMP**	**RAPD**	**RAMP**
Number of individuals	16	16	16	16
Number of population	4	4	4	4
Number of polymorphic bands	46	48	98	46
Percentage of polymorphism (%)	74.2	72.7	81.7	66.7
Nei genetic diversity (H)	0.24	0.25	0.28	0.20
Shannon information index (I)	0.36	0.37	0.42	0.31
Total gene diversity among populations (H_T)	0.24	0.22	0.27	0.19
Diversity within population (H_S)	0.15	0.12	0.18	0.12
Estimation of gene flow (N_m)	0.80	0.66	0.95	0.89
Genetic differentiation (G_{ST})	0.38	0.40	0.32	0.36
Jost genetic differentiation (D)	0.14	0.15	0.15	0.10
Mantel r test (probability)	-0.14 (P=0.48)	0.22 (P=0.12)	0.55 (P=0.03)	0.51 (P=0.05)

b. Genetic differentiation and structure

There was significant differentiation among the populations of both species. The coefficient of genetic differentiation between populations (G_{ST}) which was estimated by partitioning of the total gene diversity based on RAPD data and RAMP data was 0.38 and

0.44 in *C. fugax* and 0.32 and 0.36 in *C. hindsii*, respectively. Also, The moderate level (not low) of gene flow between populations (Nm) and Jost genetic differentiation (D) were estimated in both species between populations by RAPD data and RAMP data, indicated that significant genetic differences were present in individuals. Mantel statistic (r) was used but not any correlation between geographic and genetic distance was found at both species level and population level of *C. fugax*. In contrast, both RAPD and RAMP data predicted a significant correlation between genetic variation and geographic distance between four populations of *C. hindsii* (Table 4.8).

In *C. fugax*, multi-dimensional scaling (MDS) and multi-dimensional scaling **(**PCA) are consistent amongst RAPD and RAMP data, and clearly show the further genetic distance between population 1 and 3,4 (Figure 4.2); however, the MDS analysis was not as clear as was PCA analysis. Cluster analysis (dendrograms Figure 4.3) tend to support the structure and genetic diversity between the populations.

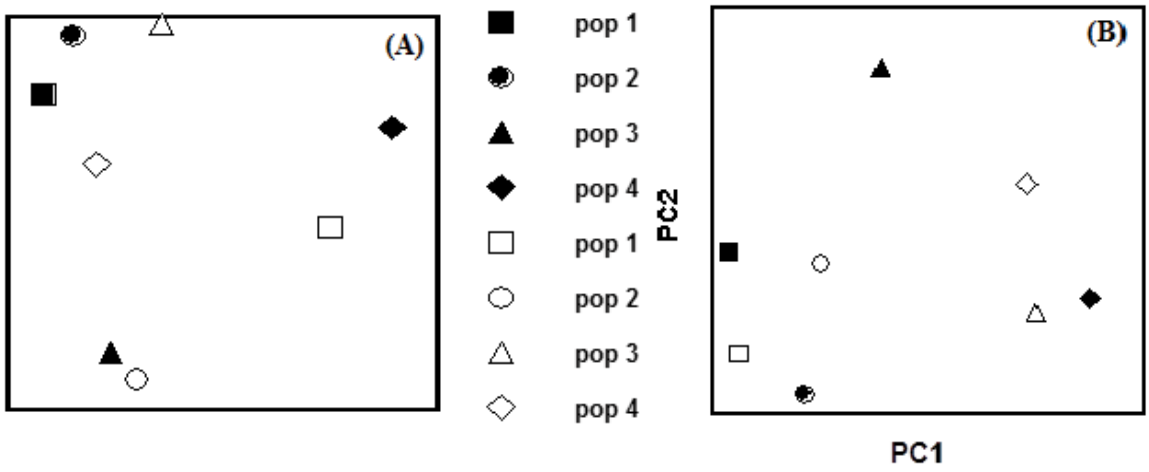


Figure 4.2 Multidimensional scaling (MDS) - (Figure A) and Principle component analysis (PCA)– (Figure B) of *C. fugax* using RAPD-PCR data (closed symbols) and RAMP-PCR data (open symbols). Where Pop1 = LV1, pop2 = LV2, pop 3 = TV, and pop4 = NF.

The UPGMA dendrogram resulting from the presence and absence of bands generated by RAPD and RAMP primers indicates a variance in the relationship between four populations LV1, LV2, NF, and TV. By RAPD data, LV2 separated itself into a single group, while NF and TV formed a close group of two, and likely expanded to secondary group with LV1. Unlikely, the results produced by RAMP-PCR showed two sets of the close genetic relationship between LV1 and LV2, and NF and TV. However, NF and TV share the similar type of ecological condition which might contribute to their genetic similarity.

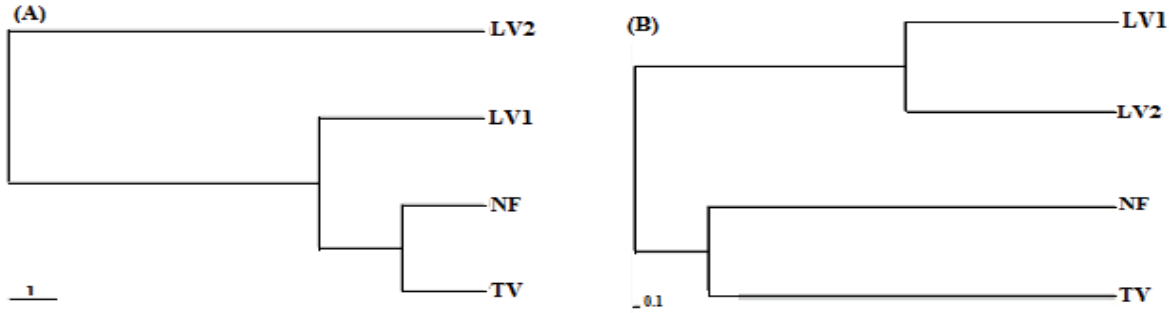


Figure 4.3 Dendrograms (cluster analysis using Nei genetic diversity) in RAPD-PCR (A) and RAMP-PCR (B) of the rare and endangered plant *C. fugax*. The result was analysed by Popgen and Treeview program.

The reason for the separation of LV2 population is probably due to the highest polymorphism which was resulted from high human-induced migration rate of its individuals as they were planted for decoration or specimen reservation purposes in home garden. That means they had been probably collected from anywhere both the wild and other gardens. NF is the furthest distance (6.5km) from TV; however, they were grouped to together as the closest genotypes. This correlation is not able to be explained by geographic distance, but these two populations share the similar natural ecological condition in common in which all individuals of TV population were originated and developed stably for an extended time in a reserved garden which shares soil type, vegetation and terrain conditions in common. Also, LV1 population which is close to the group of NF and TV may be due to their almost identical ecological conditions and origin from natural forest.

In *C. hindsii*, MDS and PCA were carried out for further population differentiation of this species and showed significant differentiation between populations. It can be seen in Figure 4.4 that MDS and PCA are consistent amongst RAPD and RAMP data, and clearly show the further genetic distance between population 1,2 and 3,4.

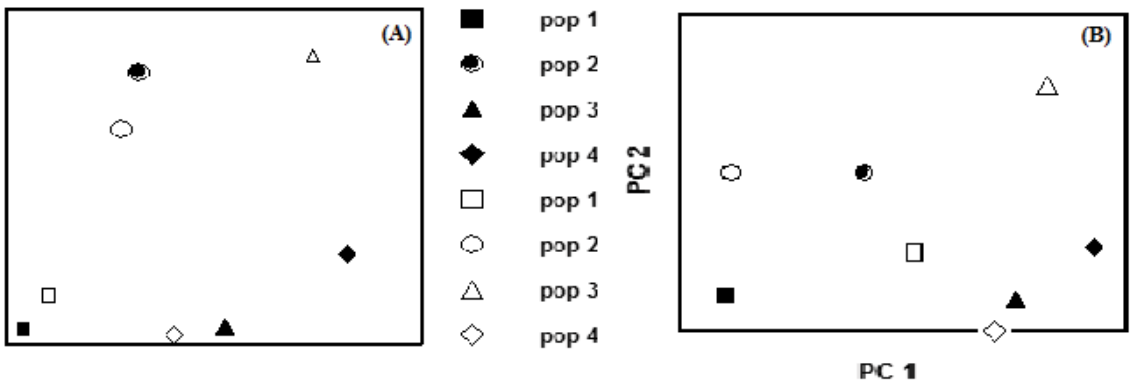


Figure 4.4 Multi-dimensional scaling (MDS) – (Figure A) and Principle component analysis (PCA) - (Figure B) of *C. hindsii* using RAPD-PCR data (closed symbols) and RAMP-PCR data (open symbols). Where Pop1 = PT, pop2 = HN, pop 3 = HB, and pop4 = HP.

Cluster analysis (dendrograms Figure 4.5) tended to support the structure and genetic diversity between the populations and showed a strong consistency between RAPD and RAMP data. The first group includes HP, and the second group includes HN and HB in one subgroup, whereas PT is separated into another subgroup. HN, HB, and PT are three closer populations, about 30 km (between HN and PT or HN and HB) to 60 km (between HB and PT), while HP is situated in Cat Ba island which is isolated with three other populations (from 150 km to 200 km). These geographic distances are relatively consistent with their relationship in the dendrogram.

The dendrogram based on Nei (1972) genetic distance produced by RAPD and RAMP data showed a similar grouping, indicating the similar effectiveness of the two markers in differentiating populations of this species. It is consistent with geographic distance between populations the distance HP to other populations (from 150km to 200km) and close relationship between HN and HB.

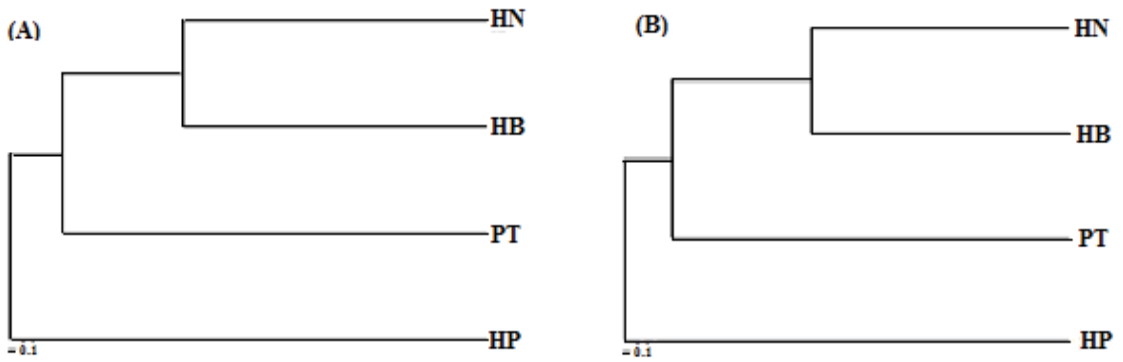


Figure 4.5 Dendrograms (cluster analysis using Nei genetic diversity) in RAPD-PCR (A) and RAMP-PCR (B) of the medicinal plant *C. hindsii*. The result was analysed by Popgen and Treeview program.

4.4 Discussion

4.4.1 DNA extraction

With two different plant species in this research, we successfully isolated high quantity and pure quality of DNA from leaves materials. In the present work, the CTAB protocol developed by Doyle & Doyle (1987) was modified in such a way that leaves from a medicinal plant of *C. hindsii* and fibrous leaves of *C. fugax* both gave good DNA yield and quality. Doyle & Doyle (1987) used 0.2% β-mercaptoethanol and not use polyvinylpyrrolidone, but in present protocol, 1% β-mercaptoethanol and 1% (w/v) polyvinylpyrrolidone PV-40 (PVP) was used in extraction buffer while still retained other steps. The increase in β-mercaptoethanol concentration and added PVP in buffer make

the extraction of DNA easier with less degradation of DNA and increase the denaturation of proteins, and remove polyphenols (Mathew et al. 2014).

Leaves of the two species are believed to contain high amounts of polysaccharides which certainly can be assessed by their fibrous and hard leaves, especially in *C. fugax*. There has been a variety of methods used to remove polysaccharides during DNA isolation procedure such as using high salt, hydrolytic enzymes (pectinase to digest pectin-like contaminants) and simply to dilute the DNA stock or/and increase times of washing the homogenate with extraction buffer. However, the most common application is the use of high concentration (more than 0.5 M) NaCl which can increase their solubility in ethanol. The current study used NaCl in combination with CTAB proved more effective results which were also reported by Murray & Thompson (1980) and Syamkumar et al. (2003). However, the current study remained the concentration of high concentration of NaCl (1.4 M) as recommended by Doyle and Doyle (1987) but not increase because several cases reported that a high amount of chemical and reagents can hinder in extraction procedure and may also cause degradation of DNA (Tiwari et al. 2012). Applying these modifications in the DNA isolation protocol, good yield and pure DNA from leaves of two selected species were obtained.

The use of modified CTAB method for DNA extraction from fibrous and mature leaf tissues of *C. fugax* and leaves of medicinal plant *C. hindsii* with high levels of secondary metabolites and polysaccharides in this research was effective, but variable. This was also reported in the study of several endangered forest tree species by (Hung et al. 2011). The average yield at about 2,000.00 (µg/g tissue) obtained from young dry leaf tissue for *C. hindsii* (ranged from 530 µg/g to 4915 µg/g) and also from mature leaves of *C. fugax* (ranged from 685 µg/g to 4195 µg/g) was much higher than those from leaves of other medicinal plants using modified CTAB protocols with PVP and high concentration of β-mercaptoethanol. For instance, Dehestani & Kazemi Tabar (2007) isolated DNA of plants with high secondary metabolites and yielded only 100-250 µg/g fresh tissues of plants. A number of other studies also used their modified CTAB protocol with different adjustment to remove contaminants from a range of medicinal plants such as increasing NaCl (to remove polysaccharides, β-mercaptoethanol and EDTA in extraction buffer, and increase incubation time with higher temperature, extended RNase treatment or even used phenol-chloroform extraction to gain good yield of DNA buffer (Murray & Thompson 1980; Rogers & Bendich 1985; Stewart & Via 1993; Porebski et al. 1997; Kumar 1999; Pirttilä et al. 2012; Mathew et al. 2014). Also, the CTAB protocol was modified and successfully used in some genetic studies on cycads species (Xiao et al. 2004; Sangin et al. 2006; Xiao & Gong 2006; Gangopadhyay et al. 2007; Yang et al. 2008; Radha et al. 2015).

The current modified CTAB protocol of DNA extraction not only isolated high yield of DNA but also produced relatively clean DNA from young dry leaves of *C. hindsii* (A_{260}:A_{280} ratio = 1.605±0.016) and pure DNA from fibrous mature leaves of *C. fugax* (A_{260}:A_{280} ratio = 1.841±0.027). However, in old dry leaves of *C. hindsii*, the A_{260}:A_{280}

ratio was only 1.098 ±0.027 and A_{260}:A_{230} ratio = 1.121 ± 0.161, indicating contaminated DNA. Another study used the modified CTAB protocol of DNA isolation from dry leaves (the same drying method using silica gel in plastic bags in this study) of related medicinal plant *Celastrus paniculatus* resulted in lower A_{260}:A_{280} ratio (between 1.2 – 1.4) compared to the young dry leaf materials of *C. hindsii.* However, the PCR products were still well-separated with high molecular weight bands without smearing (Prasad 2007). The relatively low in the purity of DNA for the old leaf of *C. hindsii* and leaves of *Celastrus paniculatus* may be due to the same type of plant materials was used and with likely similar leaf characteristics such as a high number of secondary metabolites. Unlikely, the higher yield and purer DNA from *C. fugax* leaves which are even more mature and fibrous, probably due to the nuclear DNA was better protected by cellulose cell membranes. The fibrous structure of cell wall was certainly removed by high-speed centrifuge during extraction and the use of high salt NaCl. Also, leaves of *C. fugax* are believed to contain much less number of secondary metabolites compared to other medicinal plants such as *C. hindsii*. These were probably the reasons for the purer DNA obtained from leaves of *C. fugax*. Nevertheless, the current study also reported that even within the one species, the yield of DNA could differ among sampled individuals. That is because all the leaves are slightly different regarding the time of collection, the age of leaf and the availability of material at the time of collection (particularly for critically rare *C. fugax*). The purity of DNA extracted from old dry leaves of *C. hindsii* was probably improved by adjusting the concentration of several reagents added during extraction steps, however, due to time and resources limitation, and sufficient DNA yield already obtained, the study only implemented appropriate dilution which proportionally also decreases the number of contaminants. The efficiency of DNA dilution was proven by reproducible bands separated by gel electrophoresis.

Regarding the age of the material, the yield and quality of DNA from different development stages of plant tissue was highly distinguished in *C. hindsii*. Younger tissues usually provide a higher DNA yield as well as the quality in comparison to older tissues (Porebski et al. 1997). In this study, young leaves of *C. hindsii* yielded much higher and purer DNA compared to the old leaves. The reason for this is due to young and unexpanded leaves tissues have higher cell density and less secondary structural compounds which usually formed along with tissue growth and development. Therefore, young leaves are usually superior material for DNA extraction than older leaves which have a lower number of dividing cells (Tiwari et al. 2012). In case of *C. fugax*, due to a restricted number of sampled individuals, most of the leaf materials from this plant were old, but their green state was stably conserved even through desiccated condition during transportation and storage. That would result in very high yield and purity of DNA in this species.

It is stated in literature that pure DNA should have a A_{260}:A_{280} ratio of 1.8 – 2.0, and if any results are outside this range would be considered as contaminated with proteins (<1.8) or single-stranded DNA, but more likely significant amounts of RNA if the ratio is above 2.0 (pure RNA has an A_{260}:A_{280} ratio of 2.0) (Tan & Yiap 2009). The

contaminated DNA is also identified by an A_{260}:A_{230} ratio which indicates the presence of polysaccharide and other colloidal contamination. DNA from young and old leaf tissues of *C. hindsii* had an A_{260}:A_{280} ratio of lower than 1.8 (1.61 and 1.10, respectively) which indicated excessive amounts of contaminated compounds, most likely protein or carry-over, especially from the mature leaves. The highly variable results of the A_{260}:A_{230} ratios with an average of 1.12 (on young leaves) and 1.78 (on old leaves) are also the indication of contaminant presence. To prevent the inferences of contamination to PCR amplification, all samples were treated by at least ten times dilution but still ensure to provide sufficient DNA amounts. By DNA dilution, contaminants would also be diluted by this factor before using the DNA templates for proper downstream applications. By the same method of dilution, DNA from *C. fugax* with an A_{260}:A_{280} ratio of 1.84 and an A_{260}:A_{230} of 2.1 indicating pure DNA still needed dilution as high yield was obtained (about 2000µg/g A.D. tissue).

DNA from both *C. fugax* and *C. hindsii* extracted by the same method using CTAB buffer were far too high for DNA analysis by PCR amplification, particularly RAPD-PCR and RAMP-PCR. Also, DNA yield extracted from any given extraction from two species was properly homogeneous, although there were variations among them. As a result, CTAB method used in this study is strongly recommended as the method of choice for DNA extraction for other fibrous plant species and medicinal plants containing a high number of secondary metabolites and polysaccharides.

4.4.2 PCR optimisation

a. RAPD-PCR

RAPD-PCR has been the method of choice for analysing genetic variation produced through the random amplification of genomic DNA employing short primers. With the high capability of finding homologous sequences, RAPD primers have become informative as able to scan numerous loci for polymorphisms and variations in different parts of the genome (Bardakci 2001; Chalmers et al. 1994; Al-Rawashdeh 2011; Adhikari et al. 2013; Desai et al. 2015). Subsequently, an analysis of these variations can be performed by scoring as absence or presence in these segments of the DNA targeted. Due to the simplicity, efficiency, rapidity and not require sequencing information about region amplified, RAPD primers have been selected as the informative marker to detect genetic changes only within the same species or at best very closely related species (Ren et al. 2003b). Not only has the efficient method to assay polymorphisms, but RAPD-PCR been also proven to be operative for a variety of purposes in endangered and medicinal plant genetic analysis (Al-Rawashdeh 2011; Kazemi & Hajizadeh 2012; Tahira & Saleem 2013; Bishoyi et al. 2016; Baruah et al. 2017). As a result, RAPD-PCR was selected as the key technique to detect genetic diversity and genetic differences between individuals and populations of a critically endangered species *C. fugax* distributed locally in a small and isolated area and medicinal plant *C. hindsii* which hardly found in the wild but widely cultivated. The use of RAPD-PCR for such purpose has also been reported in other plants during a long period of time such as endangered medicinal plant *Celastrus paniculatus*

(Senapati et al. 2013), medicinal and commercial plant of *Cymbopogon* (Bishoyi et al. 2016; Baruah et al. 2017), several endangered forest plants (Hung et al. 2011). However, since the nature of RAPD-PCR was recognised and reported as overestimating genetic differences (Hung et al. 2011), caution and consideration should be taken while analysing the RAPD data. In this analysis on *C. hindsii*, RAPD-PCR appeared to overestimate the genetic differences within and between populations as a considerable number of polymorphic bands difference were detected. For instance, primer OPA-7 produced a total of 32 bands, of which 30 bands (93.75%) were polymorphic. This was probably due to the genomic DNA used possesses an unusually high frequency of annealing sites for this primer (Virk et al. 1995). The results are in line with RAPD primers used to study genetic diversity of *Celastrus paniculatus* (Prasad 2007).

RAPD markers predicted an elevated level of differentiation among populations (81.67% and 74.19% polymorphism) but low intra-population diversity (from 30% to 53.33% and from 35.48% to 46.77% polymorphism) for *C. hindsii* and *C. fugax*, respectively. The population differentiation of both species was among a high level group of variation observed in previous RAPD using similar materials and closely related species. Prasad (2007) reported the use of 14 RAPD primers which generated 143 bands with 91% being polymorphic among ten accessions of *Celastrus paniculatus*. High polymorphism also exhibited in other medicinal plants using the same technique such as 88.62% in *Cymbopogon* species produced by 20 RAPD primers (Baruah et al. 2017); 86% polymorphism detected in *Clinacanthus nutans* using 17 RAPD primers (Ismail et al. 2016). However, in other different species, RAPD primers reported a lower percentage of polymorphism. There was only 10% polymorphism scanned by 16 RAPD primers in two rice mutants (M2-2, and M2-4) (Hoang et al. 2009). Nine RAPD primers generated 63% polymorphic amplicons in the genetic analysis of nine different varieties of *Cymbopogon* species (Bishoyi et al. 2016). The low intra-population diversity but high differentiated among the population of two species in this study reflected the occurrence of inbreeding. Also, while high genetic differentiation between populations of *C. hindsii* may result from wider geographic range, small population size, as well as isolation for extended periods of time which usually accompanied by inbreeding and genetic drift, were likely to be appropriate in the same situation of *C. fugax.*

Although RAPD-PCR has been proved the technique of choice to study the correlation between and within species and populations, recent biological studies recommended further research by different methods should be considered to use to identify the relationships between genetic diversity and ecological factors (Zhao et al. 2013; Zhang et al. 2005).

b. RAMP-PCR

The introduction and application of RAMP-PCR technique have been originated from SSR-PCR and RAPD-PCR in which the drawbacks of these two earlier methods were compensated. RAMP primers were developed by combining RAPD and microsatellites primers and have been proved to be more reliable in relatedness analysis with regards to

efficiency, simplicity of performance and especially suitable for plant species which have ambiguous genetic background (Sánchez de la Hoz et al. 1996; Zhang et al. 2005; Min et al. 2008). As a result, RAMP-PCR technique has been used widely to evaluate genetic diversity for a range of plant species (Fu et al. 2013; Rasouli & Karimi 2015; Long et al. 2015; Ismail et al. 2016), and also was proven superior to RAPD technique in this study. RAMP primers generated more polymorphic bands than respective RAPD primer in *C. fugax* and proved to be a valuable molecular marker (Saleh 2015). This finding concurs with the reports by Hung, De Filippis & Buckney (2011) and Ismail et al. (2016). The better effectiveness of RAMP markers because these markers are selective in their amplification which amplified conserved regions present between the microsatellites repeat sequences, whereas RAPD markers are not selective as they amplify any regions within the entire genome (Baruah et al. 2017).

In case of *C. hindsii*, in contrast, polymorphic bands generated by RAPM primers were less than that by RAPD primer alone. For example, average RAMP primers polymorphic bands detected were only 20.33, whereas that number of respective RAPD primers was 13.75. This result, however, is consistent with Hung, De Filippis & Buckney (2011) studied on endangered forest species *Sinocalamus mucclure* in which the number of band and polymorphism detected were less than those using RAPD primers alone. The explanation for this inconsistency would be that lower annealing temperature used in RAPD-PCR may induce miss-priming (Appendix 1). This is because of the nature of RAPD-primers which is to bind to sections of DNA even that do not completely match the primer sequence at a lower temperature (Birt & Baker 2000). Therefore, RAPD primers alone were able to generate more bands than their combinations with microsatellites. The contribution to the detection of polymorphisms while using RAPM-PCR technique is not only associated with microsatellites as suggested by Hayden & Sharp (2001) but may also be contributed by RAPD primers. As a result, the selection of appropriate primer combinations for those species would be more important as suggested by Min et al. (2008).

Like RAPD primer, polymorphism generated by RAMP primers was lower but still supportive to distinguish populations of two species with 73.73% and 66.67% polymorphism and remained low intra-population diversity (from 28.99% to 43.48% and from 26.15% – 43.94%) for *C. hindsii* and *C. fugax*, respectively. The use of RAMP primers in these two species agreed with RAPD primers to reveal inbreeding occurrence within these population.

4.4.3 Intra-population genetic diversity

a. Genetic diversity

Based on the percentage of polymorphism, Nei's gene diversity and Shannon's diversity index predicted by RAPD data and RAMP data, LV2 and NF, NF and LV1 respectively were the first two highest levels of diversity among populations of *C. fugax*. However, by the average results, NF exhibited the highest degree of variability while TV exhibited the

lowest level compared to other populations. NF is located in the natural forest where the population dynamic was probably the least affected by biotic and abiotic stresses. Therefore it is likely to maintain the original diversity which has been established and evolved through a lengthy period of times. Furthermore, in natural conditions of vegetation, animals and insects are supportive pollinators which help to increase genetic variation within this population. Unlikely, TV included individuals who have been developed from the smaller original population through inbreeding over generations in an isolated habitat. This condition may lead to the least genetic diversity in this population. LV2 predicted by RAPD data as the highest genetic diversity population which is probably due to overestimation of RAPD primers, and in truth, LV2 included multiple collections from diverse wild and cultivated population over the years, facilitated by horticultural trades and exchanges of the locals.

C. hindsii has a much larger range of distribution with high geographic distance and genetic distance between populations. PT of *C. hindsii* showed the highest averaged level of genetic diversity indices of RAPD and RAMP data, while the lowest was recorded in HB. PT includes multiple collections from diverse unknown populations which are probably either the wild or cultivated ones. Moreover, two out of four sampled individuals from PT were morphologically distinguished from others as their small leaf characteristics. Whereas, population HB showed the lowest genetic diversity as it was situated on the buffer zone of Ba Vi National Park which might indicate the genetic migration from vegetative introduction between home gardens within a restricted area.

b. Low genetic diversity within populations

The results of the present study using RAPD and RAMP markers revealed a low level of genetic diversity within the population and remarkable genetic differentiation among populations in both *C. fugax* and *C. hindsii*. According to Hamrick & Godt (1996), genetic diversity among populations in a species is influenced by several evolutionary factors such as mating system, gene flow and seed dispersal, geographic range as well as natural selection. One of the most influenced factors contributing to the levels of genetic diversity within a population is likely the geographic range.

C. fugax is a critically endangered cycad species with extremely narrow distribution (within several km^2). It had generally been reported that species with small geographic range tend to maintain less genetic diversity than that of geographically widespread species (Hamrick & Godt 1990). The results from this study on *C. fugax* were in agreement with this assumption as low level of genetic diversity within four populations, and high genetic differentiation between populations was estimated. Genetic diversity between at the species level and at the population level were considerably different and presented that much more genetic polymorphisms existed within species rather than within populations. These indices, although, were higher than the average of those in *Cycas guizhouensis* (Xiao et al. 2004), and *Cycas balansae* complex (Xiao & Gong 2006), they still reflected the similar pattern of genetic variation in cycads as hypothesised by Walters & Decker-Walters (1991). The hypothesis that low within-population genetic

variation with relatively high between population genetic differentiation identified as a biological and evolutionary characteristic of cycads has also been verified by other researchers Ellstrand, Ornduff & Clegg (1990) studied on *Macrozamia communis* with H = 0.08, G_{ST} = 0.27; Walters & Decker-Walters (1991) studied on *Zamia pumila* with H = 0.05; Yang & Meerow (1996) studied on *Cycas pectinata* with H = 0.13, G_{ST} = 0.39, Keppel, Lee & Hodgskiss (2002) studied on *Cycas seemannii* with H = 0.14. The low genetic diversity has been attributed to genetic drift and inbreeding effects due to their small and isolated populations for a prolonged period of times. Also, their dispersal ability is likely limited by their large and heavy seeds and pollen dispersal by insects. These biological characteristics are prone to elevated rates of inbreeding as promoting mating between individuals in close vicinity within the population (Keppel et al. 2002). The Mantel test was supportive for this conclusion as no significant correlation between genetic distance, and geographic distance was confirmed.

In case of medicinal plant *C. hindsii* which is rarely distributed in the wild, but widely cultivated in home gardens and nurseries for commercial and research purposes, has a broad range of distribution. The only wild population was found in the isolated Cat Ba National Park situated in Cat Ba Island, Hai Phong City (HP), about from 150 km to 200 km away from the other three populations investigated in this study. From the best of our knowledge, the occurrence of this species in HP has not ever been reported in the literature. Based on RAPD and RAMP data, genetic polymorphisms of this species also existed more between populations than that within populations. Genetic diversity levels represented considerably different between at species level and at population level. The lower genetic diversity within the population of *C. hindsii*, particularly estimated by RAMP data was probably due to inbreeding of isolated populations or the effect of vegetative propagation methods. Also, *C. hindsii* is an easily regenerated species by cuttings or clonal propagation which produces multiple individuals with high genetic similarity in the absence of genetic variation events (Schoettle et al. 2012). Cultivated populations of PT, HB, and HN are built mostly from home gardens, organised nurseries farms for biomass trades. All these reasons are likely to make the genetic diversity low within four population of *C. hindsii*. The result is in contrast to *Celastrus paniculatus* species in the Celastraceae family, which revealed high genetic diversity within a population due to a high level of outcrossing rates contributed by cross-pollination (Prasad 2007). However, this low level of genetic diversity is in line with Ismail et al. (2016) who studied on the medicinal plant of *Clinacanthus nutans* and reported that plant populations cultivated in the farm that practised intensive harvesting practices preventing the production of flowers which is essential for cross-pollination.

4.4.4 Inter-population genetic diversity

a. High population differentiation

Genetic diversity of plant species diverges noticeably by the mean of genetic variation distribution between populations. The present study determined that distribution through estimating Nei's genetic diversity and Shannon's diversity index which predicted the

equivalent results about the genetic structure of both species. The coefficient of gene differentiation (G_{ST}) among populations demonstrated that both species tested contained high gene differentiation as these values ranged from 0.38 to 0.40 for *C. fugax* and from 0.32 to 0.36 for *C. hindsii* (higher G_{ST} was obtained from RAMP-PCR data for both species). The high G_{ST} indicated that high degree of genetic variability was distributed among populations of these two species. Results from other species reported the different level of genetic differentiation at species level based on allozyme analysis, RAPD, SSR and RFLP markers. For instance, only 10% genetic variability (G_{ST} = 0.08) was distributed among three population of *Cycas guihouensis* based on allozyme analysis (Yang & Meerow 1999) while similar that level (G_{ST} = 0.39) was estimated in *Cycas pectinata* (Yang & Meerow 1996) and much higher among populations of *Cycas balansae* complex (G_{ST} = 0.60). These different results showed that genetic differentiation of a species greatly depends on species, analysis method, and conservation status.

While the low genetic diversity within population was mainly due to inbreeding and genetic drift, the high population differentiation in *C. fugax* was probably as the consequence of the long-term evolutionary history of the species such as shifts in distribution, habitat fragmentation and population isolation, as well as mutation, genetic drift, mating systems, gene flow and natural selection (Schaal et al. 1998; Xiao et al. 2004). Of those possibilities, genetic drift and mating systems are likely the major causes for highly differentiated populations of *C. fugax*. As mentioned above genetic drift, on the one hand, reduces variation between individuals of populations, but on the other hand, directly increase differentiation between isolated populations (Xiao et al. 2004). As a result, genetic variation usually changes along with genetic drift. *C. fugax* has relatively smaller cream colored male cones (10-12 cm, 2.5 – 4cm in diameter) and larger ovoid seed (roughly 3cm x 2cm) than other cycads (Osborne et al. 2010). These reproductive characteristics limit its dispersal capability by decrease efficiency of pollinators and seed carriers; directly contributing to low level of gene flow and high level of differentiation. In addition, as estimated by Yang & Meerow (1996), the gene flow between cycads was only estimated to be 2-7km. In this study, the distance from TV population to LV 1 and NF are from 8 km to 9 km. Therefore, it is likely that gene exchange between TV and other two populations is restricted.

Unlikely, the high population differentiation of *C. hindsii* is due to a recently rapid reduction of population size with insufficient time to isolate and spread gene flow (N_m) (Maguire & Sedgley 1997). Moreover, geographic distance is also a major factor that contributes to high genetic diversity at the species level. *C. hindsii* has been widely distributed in Vietnam, occurring in wild and cultivated habitat with diverse climatic, geographical and edaphic conditions (Muoi et al. 2009; Nam et al. 2013). These differences have an impact on genetic diversity of a species. Genetic drift over thousands of generations would lead to significant divergence among populations. The results from RAPD and RAMP analyses indicate that genetic drift might have occurred among the studied populations of *C. hindsii*, additional to increasing vegetative propagation activities, thereby producing differentiation. Furthermore, *C. hindsii* is a high cross-

pollinated plant species, an introduction of a different variant, especially with the occurrence of genomic mutations, certainly, increase population differentiation. Pollen flow and local selection may also facilitate this trend. The higher genetic diversity of *C. hindsii* populations confirmed by the UPGMA dendrogram, (which were consistent with RAPD and RAMP data) where four populations were classified into three different groups and sub-groups, and well-separated to each by MDS and PCA analyses.

High population differentiation was not only consistently predicted by Nei's genetic diversity and Shannon's diversity index but also the coefficient of gene differentiation and Jost genetic differentiation (D). G_{ST} among populations demonstrated that *C. fugax* contained high gene differentiation as these values ranged from 0.38 to 0.40 (higher G_{ST} was obtained from RAMP-PCR data), which indicated that maximum 40.00% of the genetic variability was distributed among populations of this species. According to Zhao et al. (2016), gene differentiation (G_{ST}) values of $0 \leq 0.05$ is defined as low, $0.05 \leq 0.15$ as medium and $0.15 \leq 0.25$ as high level of genetic differentiation. Also, according to Baruah et al. (2017) who had classified N_m values < 1 as low, N_m > one as moderate and $N_m > 4$ as extensive. In this regard, the average N_m value of 0.80 and 0.66 by RAPD and RAMP, respectively in *C. fugax* were at a low level (less than one migrant per generation into a population). Results from other species reported the different level of genetic differentiation at species level based on allozyme analysis, RAPD, SSR and RFLP markers. For instance, only 10% genetic variability (G_{ST} 0.08) was distributed among three populations of *Cycas guihouensis* based on allozyme analysis (Yang & Meerow 1999), while higher levels (G_{ST} 0.39) was estimated for *Cycas pectinata* (Yang & Meerow 1996), and *Cycas guihouensis* (G_{ST} 0.43, N_m 0.33) (Xiao et al. 2004); higher levels (close to our values) were recorded for *Cycas balansae* (G_{ST} 0.60) (Xiao & Gong 2006). These values demonstrate that genetic differentiation of a species in small populations greatly depends on the geographic location, distance and analytical methods used.

For *C. hindsii*, the G_{ST} value is 0.32 and 0.36 (by RAPD and RAMP, respectively), indicating a high level of gene differentiation, while the average N_m value of 0.95 by RAPD and 0.89 by RAMP were only from low to moderate levels. These results are similar to a study on a medicinal plant species *Justicia adhatoda* with $G_{ST} = 0.30$ and $N_m = 1.28$ (by RAPD) also indicating moderate N_m and high level of G_{ST} values (Kumar et al. 2014). High G_{ST} values mean high variations within populations, whereas moderate N_m values estimate that one species in every generation can link the gene pools among populations. In summary, results from the molecular data showed that a considerable amount of genetic variation between populations was present at the fragmented sites sampled. Similar results were found several studies, such as cycads (Xiao et al. 2006); *Flemingia macrophylla* (Heider et a.l 2007); and some rare and endangered forest species (Hung et al. 2011).

b. Populations relationship

Genetic structure of *C. fugax* was determined using different parameters, where the values were found to be higher for RAPD markers than that from RAMP markers. MDS, PCA,

and dendrogram were constructed to explore the relationship among populations based on Nei's genetic distance and Bray-Curtis cluster values. The group (population) divisions were consistent between RAPD and RAMP data and showed the further genetic distance between the populations. Similar findings were observed by Hung et al. (2011), where clusters show a clear separation between populations using RAPD and RAMP. Also, low genetic variation was found in all population, therefore, this species should benefit from the introduction of genetic diversity by breeding and hybridisation. There was not any relationship between genetic differentiation and geographic distance among *C. fugax* populations, suggesting the influence of other factors, such as human-induced interference and outbreeding barriers to the genetic differentiation.

In contrast, geographic distance between populations of *C. hindsii* were reflected how they were grouped in dendrogram as predicted my Mantel Test ($r = 0.55$, $P = 0.03$ by RAPD data, and $r = 0.51$, $P = 0.05$ by RAMP data). This population differentiation is not only due to geographic distribution but also induced by inbreeding occurrence and the expanding of vegetative production via cutting and clonal propagation in most of the tested populations.

4.5 Conclusion

In this study, DNA-based molecular markers revealed low polymorphisms within the population, however, high genetic differentiation among populations for both species. The underlying causes of the differences appeared to be the geographic distance among populations of *C. hindsii*, but in contrast, the differentiation in *C. fugax* was corresponded with human and physical disturbance of the sites, i.e. in fragmentation. The study results serve as guidance for the delineation of *in situ* and *ex situ* conservation approaches, in which all four population of each species should be selected for preservation, especially *C. fugax*, as all four sites are within a very narrow geographic restriction. An urgent response for the conservation of the critically endangered species *C. fugax* is crucial, with the application of effective methods of sexual reproduction through an increase in cross-pollination. Also, natural populations of both species should be paid more attention in conservation program as they contain high genetic diversity, compared to other cultivated populations. The generated baseline data are valuable for follow-up research (e.g., re-assessment of genetic diversity of rehabilitation programs) and for future decision-making processes associated with management and conservation of genetic resources.

CHAPTER 5

ENZYMES AND METABOLITES

5.1 Introduction

Plants exposed to adverse abiotic stresses usually display a range of morphogenetic and biochemical responses to redirecting growth and metabolism. Plants may alter their morphological characteristic through a mechanism of inhibiting cell elongation, stimulating cell division, and altering cell differentiation status (Potters et al. 2007). These alterations may be reflected in common molecular processes such as increasing production of reactive oxygen species (ROS), and modifying metabolism and phytohormone transport, and signalling responses (Potters et al. 2007). The underlying mechanism is that under oxidative damage, plants undergo an array of morphological, physiological and biochemical responses to mitigate stress exposure and enhance repair of damaged systems. As a result, the stress tolerance of plants may also be boosted through plant responses, such as antioxidant defences (Mittler 2002).

In order to limit the exposure to unfavourable environmental conditions, plants have evolved a large variety of distinct morphological adaptations and biochemical mechanisms. A number of studies on plants have revealed differential morphological responses such as decreased root, shoot or leaf elongation (Amudha et al. 2005), increased leaf thickness and reduced specific leaf area (Liu et al. 2005) against stresses such as UV-B radiation, toxic heavy metals, nutrient-deficiencies and hypoxia (Pierik et al. 2006). Protective mechanisms against the action of activated oxygen species exist, and central to these mechanisms include the action of antioxidants (Kamiloglu et al. 2016). Given that plant cells and organelles are exposed to ROS, plants have evolved biological processes to prevent ROS formation and scavenge ROS by the accumulation of low molecular weight antioxidants such as ascorbic acid, glutathione, α-tocopherols, amino acids (e.g. proline), sugars, carotenoids, and quinic acid derivatives. Along with the accumulation of these anti-oxidative metabolites, various other kinds of specialised metabolites, such as flavonoids are also produced in response to abiotic stresses (Nakabayashi & Saito 2015).

C. hindsiii is a type of traditional Vietnamese medicinal plant commonly grown in Northern Vietnam in household gardens and nurseries and used as dry leaves, chopped dry stems, branches or roots (The Asian Foundation 2012). Leaf features are highlighted by broad rectangular elliptic and narrowly ovate-elliptic leaf blade (broad and narrow leaf) (Zhixiang & Funston 2008). However, these features have not yet been identified phenotypic responsive to stress, i.e. induced morphogenetic responses. In plants, phenotypic variation can be explained based on the diverse geographical origins of the variants and local human selection. As a result, only the plant part of interest is selected, and potential ecotypes and botanical varieties are likely to form (Diederichsen & Hammer 2003). The *C. hindsii* collection in Vietnam is characterised by several phenotypic variations for traits of interest, such as broad leaf blade variant in Hoa Binh province,

which has been chosen for cultivation because it provides larger biomass for trades (according to the study survey). The finding is worth to pay attention as it is consistent with the goals of *ex-situ* conservation, which is designed to preserve sufficient genetic and phenotypic diversity of important characteristics, especially for commercial medicinal plants.

Recently, studies of *C. hindsii* have mainly focused on phytochemical investigations, regardless of the biological variation of different germplasms or leaf morphological variants (Yao-Haur Kuo et al. 1995; Huang et al. 2000; Ly et al. 2006; Hu et al. 2013; 2014). Therefore, the relationship between morphological characteristics and active component content has already been noticed and studied (Khanna et al. 2014; Fu et al. 2016). Thus, it is worth investigating the mechanisms by which leaf variations affect plant ecotypes in a comparative study of secondary metabolites accumulated in leaf extracts of *C. hindsii*.

Plant also provides an important source of proteins for a healthy diet with vital amino acids that the body cannot self-produce (DRI Committee 2005). Proteins are an integral part of the chemical machinery for cell growth, division, and differentiation in young developing cells. Quantitation of proteins by LC/MS is very time consuming and costly as the complex nature of protein compounds with diverse structure and composition, while colourimetric spectrophotometer protocol is more economical to quantitate protein content of plant extracts. Also, proteases which degrade proteins by splitting internal peptide bonds, including proteolytic enzymes and endopeptidases and are produced by plant tissue under development process, senescence and stress (Roy-macauley et al. 2001). Such enzymes provide a quantitative biochemical measure of processes such as the regulation of intracellular protein breakdown and turn-over (Sanman & Bogyo 2014; Kidric, Kos & Sabotic 2014).

Metabolomics, which is designed to provide general qualitative and quantitative profiles of metabolites in organisms exposed to different conditions, enable us to monitor the spatial and temporal distribution of target phytochemicals. In fact, assigning bioactive compounds from the complex is a central challenge of natural product research. More specifically, metabolomics is enabling a better understanding of medicinal plant and the identification of important metabolic quantitative trait loci for enhanced breeding (Rai et al. 2017). The rapid development of the main techniques used in the analyses of metabolites (e.g. gas chromatography, high-performance liquid chromatography, and nuclear magnetic resonance) is increasing the application of metabolomics in many aspects of natural products discoveries (Jensen et al. 2016). The combination of bioassay-guided fractionation with untargeted metabolite profiling improves the identification of active components (Wong et al. 2014). While the modern techniques require both time and labour, they should be carried out depending on the research (Moniruzzaman et al. 2014). In this study, it is still a challenge since no comprehensive database exists up to date for secondary metabolite analysis, particularly for medicinal plants. As a result, a high throughput screening method (targeted profiling) which allows implementing a large number of samples in duplicates maybe suitable for investigation a limited number of metabolites of

interest to describe the intraspecific diversity and geographic patterns of variation (Shulaev et al. 2008). Therefore, Stratil, Klejdus & Kuban (2006) reported that using spectrophotometric measurements with specific analytical reagents is adequate to obtain sufficient information for evaluation of contents of secondary metabolites in a comparative study.

The purpose of this comparative study was to determine the total protein content, the amount of anti-oxidative metabolites including proline, glutathione, tocopherol, flavonoid, and phospholipid in two forms of leaf phenotypes: narrow leaf and broad leaf variants of *C. hindsii*.. In this study, the protein degradation of leaf material in *C. hindsii* will be examined, to understand the response of protease activities whether increasing or decreasing between narrow leaf (NL) and broad leaf (BL) cultivars. Malondialdehyde, a product of lipid peroxidation will be measured because it is also a biomarker measurement of stresses. It is expected that information from antioxidants, enzymes, and stress biomarkers will be supportive to select appropriate populations, which is better in stress tolerance and pharmaceutical properties, for commercial utilisation of medicinal plant *C. hindsii*. For *C. fugax*, although there has been evident about its pharmaceutical properties from seeds, due to materials insufficiencies (restricted by Australian quarantine regulations) and limited resources, the further phytochemical investigation will be omitted.

5.2 Materials and methods

5.2.1 Leaf material

C. hindsii leaves (two broad leaf plants and two narrow leaf plants) were collected from two adjacent gardens which are similar in soil and fertilised conditions in Phu Tho province. The specimens were classified into the two morphogenetic groups: narrow leaf and broad leaf and authenticated by Dr Pham Thanh Loan (Dean of Agroforest and Fishery Faculty, Hung Vuong University, Vietnam). The classification into small leaf plants and broad leaf plants was based on leaf length, leaf width, leaf thickness, and yield. Also, other morphological parameters of the plant were taken into consideration during classification processes such as number of primary branches, crown diameter, and plant height. All the morphological parameters are presented in Table 5.1 below.

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

Traits (unit)	NL		BL	
	Mean	**SE**	**Mean**	**SE**
Leaf length (cm)				
Leaf width (cm)				
Leaf thickness (mm)				
Leaf yield (g/m^2)				

Leaves were manually harvested from mature plants, cleaned and immediately protected away from moisture and direct sunlight in silica filled sealed plastic bags in the field,

before shipment. Air-dried leaves after transporting from Vietnam to Australia and obtaining quarantine permission were store in -80^{0}C freezer in UTS Science laboratory before use. Air dried leaf material was weighed using an electronic balance (Sartorious, Quintix224-1S, Germany, accuracy ± 0.0001 g) and sliced into thin pieces (1 - 2 mm^2) for enzyme and metabolite extraction and extracted in various specific buffers using a cold pestle and mortar (see Section 4.2.3 – 4.2.9).

Air dry weigh and protein basis were used to standardize the quantity of enzymes and metabolites. Dry weight was also measured to quantify the water content using air-dried leaf tissues in the oven at 65^{0}C overnight. The total evaporable moisture content and percentage water content (P) of tissue were calculated with the formula:
P = (ADW – DW)/ADW x 100 where ADW – Air dry weight; DW – Dry weight

5.2.2 Protein content

Protein was assayed using a commercial assay kit (Bio-Rad). The Bio-Rad protein assay is based on the differential colour change of a dye in response to various concentrations of protein. The absorbance maximum for an acidic solution of Coomassie Brilliant Blue G-250 shifts from 465 nm to 595 nm when binding of protein occurs. Lambert-Beer's-Law with the appropriate ratio of dye volume to sample concentration results in an accurate quantification of total soluble protein (Einsenthal & Danson 1992). To ensure the linearity of the Beer-Lambert Law, and avoid the deviation in absorptivity coefficients at high concentrations, the solutions were diluted with relevant dilution factors

5µl of enzyme/metabolite solution and 995 µl of the diluted dye reagent (Bio-Rad reagent diluted 1:5) were added to clean dry test tubes (2 ml) and mixed thoroughly with a vortex. After 5 min, absorbance was measured at 595 nm in glass cuvettes. The blank consisted of the diluted dye reagent. Cuvettes were rinsed and cleaned between measurements with reagent grade water followed by acetone to ensure the removal of all blue dye complex.

5.2.3 Proteases

Acid and neutral protease activities were determined by the amount of α-amino acids (α-NH_2) released from haemoglobin according to the conditions and methods described by Hampp & De Filippis (1980). 1 unit of protease enzyme activity was defined as that which released 1µmol α-NH_2. The assay mixture was composed of the following constituents:

(a) Buffer: 0.5 ml of Citric acid buffer pH 5.4 (for acid protease extraction);
(b) Buffer: 0.5 ml Phosphate buffer pH 7 (for neutral protease extraction);
(c) Substrate: 0.25 ml of 4% haemoglobin in appropriate buffer above;
(d) β-mercaptoethanol (BME): 10 mM, 0.25 ml;
(e) Enzyme solution (tissue extract): 0.5 ml.

100 mg of fresh leaf tissue was ground by mortar and pestle in two separated extraction buffers for acid protease and neutral protease. Constituents above were added and were dispensed into a 2ml Eppendorf tube then incubated at their optimum temperature (40^{0}C ± 2^{0}C) for 90 min for both enzymes (acidic and neutral) with sufficiently vigorous shaking

to keep the substrate in an even suspension. The reaction was stopped with 0.25 ml of 40% trichloroacetic acid (TCA) and allowed to stand for half an h in a refrigerator. Blanks were prepared for each treatment by adding TCA before the enzyme (i.e. tissue) solution. The reaction mixture was centrifuged at 12,000 g for 2 min, and 0.5 ml of the supernatant solution was analysed for α-amino acids (α-NH2) with the ninhydrin method. The ninhydrin reagent solution consisted of 0.008 g of stannous chloride dissolved in 5 ml of citrate buffer (pH 5) and 0.2 g of ninhydrin dissolved in 5 ml ethylene glycol. The reagent remained stable for 24 h when stored at 4oC. The reagent was kept on ice during experiments to ensure stability. 0.5 ml of the ninhydrin reagent solution was added to 0.5 ml of the incubated sample solution and heated for 20 min in a vigorously boiling water bath. The mixture was allowed to settle at room temperature for 15 min. Absorbance was measured at 570 nm using a spectrophotometer (Biochrom Libra S12 UV/Vis) with distilled water as the blank. Protease activity was calculated concerning a standard curve using L-glycine as the reference amino acid.

5.2.4 Proline

Aqueous sulphosalicylic acid was used during selective extraction. Interfering materials such as proteins are precipitated as a complex, and others were mainly removed by absorption to the protein-sulphosalicylic acid complex. The extracted proline is reacted with ninhydrin in acidic conditions to form the blue chlorophore and read at 520nm (Bates et al. 1973). The assay mixture was composed of the following constituents:

(a) Sulfosalicylic acid 3%
(b) Toluene
(c) Ninhydrin in glacial acetic acid: 1.25 g ninhydrin dissolved in 30 ml glacial acetic acid and 20 ml 6 mol/ L phosphoric acid.

100 mg of leaf samples were ground with a mortar and pestle in liquid nitrogen and extracted in 3% sulfosalicylic acid. Then mixtures were filtered and made up to 2 ml. This 2 ml of filtrate was reacted with 2ml acid ninhydrin in glacial acetic acid, and the mixture was boiled in a water bath for 1 h, then the reaction quickly terminated in ice. The reaction mixture then was extracted with 3 ml toluene followed by vortex and left at room temperature for 30 min until separation of the two phases. A toluene blue layer appeared and ready to be read by a spectrophotometer (Biochrom Libra S12 UV/Vis) at 520 nm.

5.2.5 Glutathione (GHS)

The total concentration of glutathione was determined in a kinetic reaction based on the oxidation of GHS to GSSG by the Ellman reagent DTNB (5,5'-dithiobis-(2-nitrobenzoic acid)) to produce a conjugate and TNB anion that can be detected by UV/Vis absorbance (Tietze 1969). The enzyme glutathione reductase reduces the GSSG releasing the GSH that can react with another molecule of DTNB. Therefore, instead of a single determination of how much DTNB reacts with GSH, the rate of TNB production is

measured over a set time, as that is proportional to the initial amount of GSH (Forman et al. 2009). The assay mixture was composed of the following constituents:

(a) Trichloro acetic acid (TCA) 5%,
(b) Phosphate buffer (pH8.0) 0.2 M.

100 mg of fresh leaf tissue was ground by mortal and pestle to a fine powder in liquid nitrogen and then treated in 2 ml of TCA (5%). The mixtures were centrifuged at 12,000 g for 15 min in cold ice temperature (4^0C) and the supernatant was taken. After deproteinization, the supernatant was allowed to react with 2.9 ml of Ellman's reagent (2 ml of 1 mM 5, 5'-dithiobisnitro benzoic acid in 0.9 ml of 0.2 M phosphate buffer pH 7.0). The absorbance of the yellow product was read at 412 nm and 340 nm in a spectrophotometer (Biochrom Libra S12 UV/Vis). Pure GSH was used as standard for establishing the calibration curve.

5.2.6 Tocopherol

This method used the reaction which was based on the reduction by tocopherol of ferric ions to ferrous ions, which then forms a red colour complex with α, α' – dipyridyl reagent (Kivcak & Akay 2005). The assay mixture was composed of the following constituents:

(a) Absolute ethanol
(b) Xylene (100%)
(c) H_2SO_4 0.1 N
(d) α, α' – dipyridyl reagent: 1.2gm of dipyridyl reagent was weighed and dissolved in ethanol, and the final volume was made up to 1 litter, then stored in a dark bottle in the refrigerator until use.
(e) Ferric chloride $FeCl_3$ (0.3%): dissolve 0.3 g in 100 ml of absolute ethanol.

100 mg of fresh leaf tissue was ground to a fine powder in liquid nitrogen and then treated in 2 ml of 0.1 N H_2SO_4. Then mixtures were incubated at room temperature overnight and filtered. Afterwards, extracts (1.5 ml) were transferred to a 15ml tube, then 1.5 ml methanol and 1.5 ml xylene were added. The mixture was vortexed for about 1 minute to allow separation. 1.0ml of the supernatant was collected and transfer to new tubes, then α, α' – dipyridyl (1.2gram in 100% ethanol) were added to the mixture. Next, FeCl3 (3%) was added, and solutions were kept for 1 minute. The absorbance of the solutions was read at 520 nm and 460nm using a spectrophotometer (Biochrom Libra S12 UV/Vis). During their preparation, all solutions were kept away from the light. α-tocopherol standard was dissolved in xylene in a concentration range of 0–320 µg/ml and a calibration curve was calculated.

5.2.7 Flavonoids

The vanillin reaction involves the reaction of an aromatic aldehyde, vanillin with the meta-substituted ring of flavonoids to yield a red adduct according to Swain & Hillis (1959). The assay mixture was composed of the following constituents:

(a) Vanillin: 1% dissolved in 70% sulphuric acid
(b) Absolute Methanol: Water (2:1)
(c) Absolute Methanol: Water (1:1)

100 mg of leaf samples were ground with a mortar and pestle in liquid nitrogen and extracted in 1.5 ml methanol: water (ratio 2:1). The centrifuged (12,000 g) and collected supernatant was transferred to a new tube (15ml). The solid phase was extracted again in 1.5ml of methanol: water (1:1). Centrifuged (12,000 g) the supernatant collected and combined with the first extract. The combined extract was evaporated to dryness with slight heat using an incandescence bulb lamp in the fume hood. 4ml of working vanillin reagent was added to the dry extract and boiled in water bath for 15 min. The sample was then left on a bench to cool down, and the absorbance at 340 nm was read (340 nm for flavones and flavanones) (Morina et al. 2015) by a spectrophotometer (Biochrom Libra S12 UV/Vis). The colour became dark black due to a high level of flavonoids; therefore, the solution was diluted 1:10 before reading.

5.2.8 Phospholipids

Phospholipids in lipid extracts were estimated by phosphorous release and determined through acidic digestion. The assay mixture was composed of the following constituents:

(a) Chloroform: Methanol (1:1),
(b) Concentrated acid sulphuric,
(c) Nitric acid (5N): slowly add 318.522 ml of 70% Nitric acid to 250 ml deionised water. Adjust the final volume of solution to 1000 ml with deionised water.
(d) Sodium hydroxide (1N): dissolve 50 g of AR NaOH in 1 litre distilled water.
(e) Phosphorous reagent A: Add together 76.80 g Ammonium Molybdate dissolved in 200 ml distilled water, and 1.755 g Antimony Potassium Tartrate dissolved in 100 ml distilled water. Add carefully a mixture of 500 ml of distilled water and 896 ml concentrated AR grade sulphuric acid. Make up to 2 litters of distilled water. The resulting solution should be clear and colourless, and it will keep for a long time but should be stored in a cool place. Looks for cloudiness or precipitation when it is off.
(f) Phosphorous reagent B: Dissolve 1.70 g ascorbic acid in 100 ml distilled water (this keeps only for 48 h) and add 50 ml of reagent A just before using in tests. Make up to 200 ml with distilled water just before use. This final reagent is unstable and should be prepared fresh every day.

100 mg of leaf samples were ground with a mortar and pestle in liquid nitrogen and extracted in 2 ml chloroform: methanol (1:1). The volume was evaporated to dryness with slight heat using an incandescence bulb lamp in the fume hood. 1ml sulphuric acid was added to the dry extract and a brown colour allowed developing. 2-5 drops of Nitric acid (5N) were added until a colourless solution was obtained. The samples were then added to 2 ml of distilled water. The final mixture of sample included: 2 ml sample from before, 1 ml 1N NaOH, 10 ml distilled water and 2 ml phosphorous reagent B; which was heated at 60^0C for 10 min and cooled down to room temperature for reading (at 880 nm) by a spectrophotometer (Biochrom Libra S12 UV/Vis, UK).

5.2.9 Malondialdehyde (MDA)

Malondialdehyde is one of end products formed via the breakdown of certain lipid peroxidation products. At low pH and elevated temperatures, MDA readily participates in a nucleophilic addition reaction with thiobarbituric acid (TBA), generating a red fluorescent 1:2 MDA: TBAR (thiobarbituric acid-reactive-substances) adduct. The TBA reaction method and equations were reported by Dhindsa, Plumb-dhindsa & Thorpe (1981). The assay mixture was composed of the following constituents:

(a) Trichloroacetic acid (TCA) 0.1%,
(b) Trichloroacetic acid (0.1%) + Thiobarbatate (TBARS) 0.5%.

100 mg of leaf samples were ground with a mortar and pestle in liquid nitrogen and extracted in 2 ml 0.1% (w/v) TCA. The resulting solution was transferred with a pasteur pipette to 2 ml Eppendorf tubes and centrifuged at 12,000 (g) for 5 min. 1 ml 20% (w/v) TCA containing 0.5% (w/v) thiobarbituric acid was added to 1 ml plant tissue sample in a 10 ml test tube. The reaction mixtures were vortexed thoroughly and incubated at 95^{o}C in a water bath for 30 min. The reaction was terminated by placing sample tubes in an ice bath. The mixture was centrifuged again for 10 min at 12,000 g and absorbance recorded at 532 nm and 600 nm by a spectrophotometer (Biochrom Libra S12 UV/Vis). The control blank was 1 ml 0.1% (w/v) TCA and 4 ml 20% (w/v) TCA containing 0.5% (w/v) thiobarbituric acid. One absorbance unit (532 nm) was equivalent to 155 mmol/l MDA in the sample.

5.2.10 Replications and statistics

Enzyme and metabolites experiments were designed to ensure the inclusion of a sufficient number of replicates for statistical analysis. Statistical analysis was conducted on the data using means, standard errors, one-way analysis of variance (ANOVA). Duncan Test was applied for further statistical analysis following the rejection of the null hypothesis by ANOVA. This method allowed comparison of every possible pair of means using a single level of significance (P = 0.05).

5.3 Results

5.3.1 Protein content

Water content, leaf thickness, and protein content on the basis of air-dried weight, dry weight and leaf area were determined in NL and BL of *C. hindsii*. The results are shown in Figure 5.1. Although the variants were cultivated in the same area, the difference in some parameters above were significant. There was a significant difference in the percentage of water contained in NL and BL. BL contains 42.53% water, while there was much lower water content in NL at only 15.14%. Also, the thickness, length, and width of the leaf blade in BL were all significantly different from NL (Appendix 3).

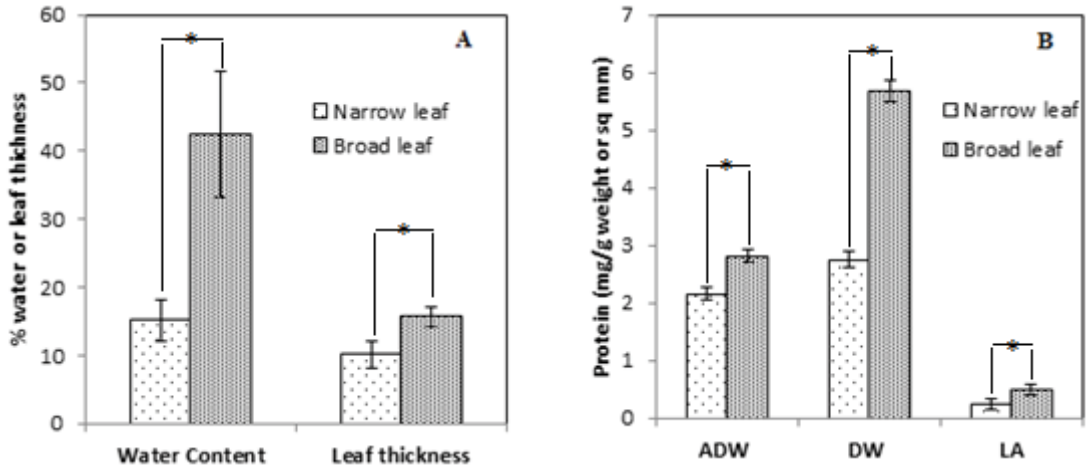


Figure 5.1 Percentage water and leaf thickness (A), protein content on the basis of air-dried weight (ADW), dry weight (DW) and leaf area (LA) (B) in narrow leaf (NL) and broad leaf (BL). An asterisk indicates that treatments were significantly different ($p<0.05$).

Protein content was shown to be statistically different between NL and BL based on all counts and from 2.16 ± 0.12 mg/g ADW to 2.82 ± 0.11 mg/g ADW; from 2.76 ± 0.14 mg/g DW to 5.68 ± 0.18 mg/g DW, and from 0.24 ± 0.1 mg/mm^2 to 0.485 ± 0.09 mg/mm^2 respectively. Of those, the protein content difference on a DW basis is very apparent in BL, which is two times higher than that in NL.

5.3.2 Protease

Overall, AP activity is much stronger than NP activity, and AP activity was significantly increased in NL compared to BL, while there was no statistically significant difference in NP activity in all leaf types (Figure 5.2). According to one-way ANOVA analysis and Duncan Test, the average rate of AP reaction significantly increased in NL average 5.45 ± 0.25 (µmol/h/g ADW) and 2.33 ± 0.11 (µmol/h/mg protein) in BL (Appendix 3). In contrast, NP activities were similar between NL and BL on both bases.

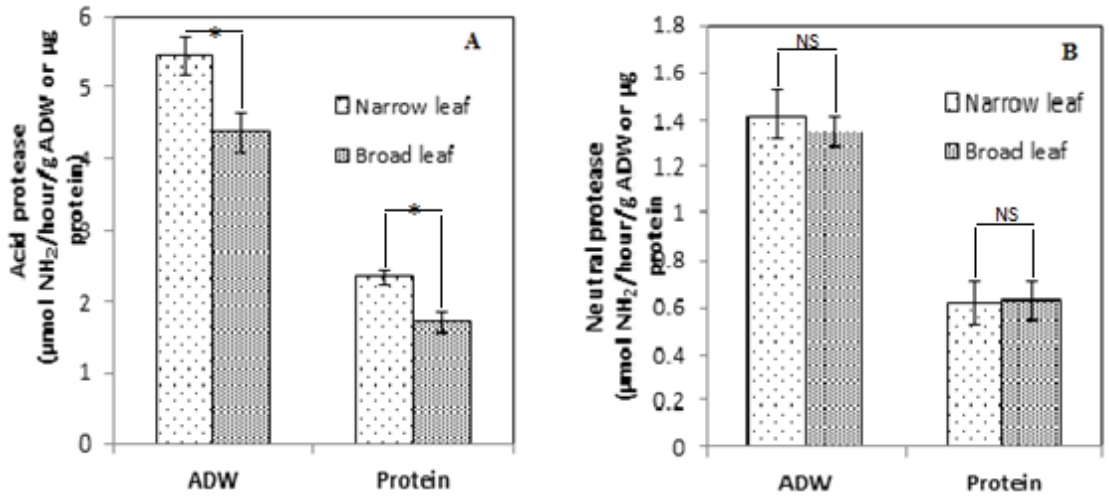


Figure 5.2 Acid protease (A) and neutral protease (B) activity on a protein basis and a dry weight basis in NL and BL extracts. Mean values ± (n=8). Data and result were based on one-way ANOVA and Duncan Test. An asterisk indicates that treatments were significantly different (p<0.05). NS indicates that treatments were not significantly different.

5.3.3 Proline and glutathione

Proline content increased significantly in NL with 12.71 ± 0.77 (µmol/g ADW) and was significantly lower in BL with 9.7 ± 0.78 (µmol/g ADW). However, there was not a statistically significant difference in proline content between NL and BL on a protein basis (Figure 5.3). Although glutathione content in BL was measured higher than that in NL on ADW basis, there was not a statistically significant difference between NL and BL extracts on both ADW and protein basis (one-way ANOVA and Duncan Test).

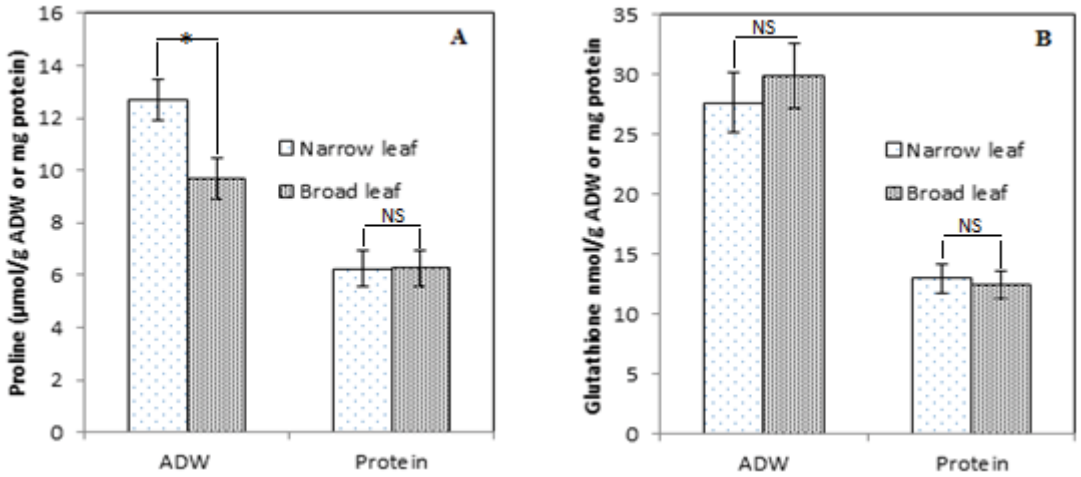


Figure 5.3 Proline (A) and glutathione (B) content on a protein basis and an ADW basis in NL and BL extracts. Mean values ± (n=8). Data and result were based on one-way ANOVA and Duncan Test. An asterisk indicates that treatments were significantly different (p<0.05). NS indicates that treatments were not significantly different.

5.3.4 Tocopherol and flavonoid

Both tocopherol and flavonoids in BL extract are significantly higher than that in NL extracts (Figure 5.4). These differences are statistically significant (one-way ANOVA and Duncan Test). The tocopherol content expressed on ADW and protein bases, varied from 18.81 ± 1.5 (µg/g ADW) or 6.77 ± 1.2 (µg/mg protein) on NL; to 22.29 ± 1.23 (µg/g ADW) or 9.79 ± 1.1 (µg/mg protein) on BL respectively. In the same way, flavonoids content increased significantly from 18.75 ± 0.96 (mg/g ADW) or 7.43 ± 0.49 (mg/mg protein) in NL to 23.16 ± 1.03 (mg/g ADW) or 8.78 ± 0.52 (mg/mg protein) in BL. From the data, both flavonoids and tocopherol concentrations are lower if measured based on protein content but still significantly different.

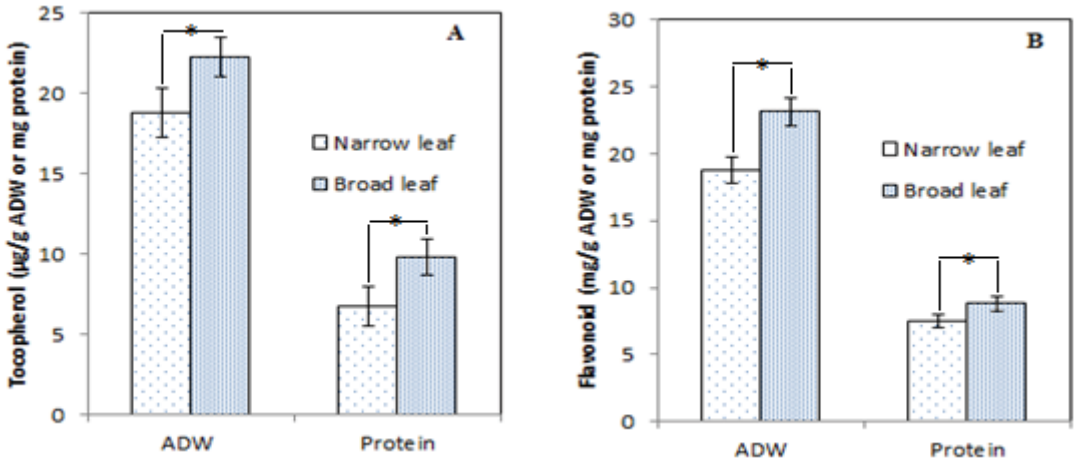


Figure 5.4 Tocopherol (A) and flavonoid (B) content on a protein basis and a dry weight basis in NL and BL extracts. Mean values ± (n=8). Data and result were based on one-way ANOVA and Duncan Test. An asterisk indicates that treatments were significantly different ($p<0.05$).

5.3.5 Phospholipid and malonyldialdehyde (MDA)

It is apparent from Figure 5.5 that phospholipid content in BL is statistically elevated on ADW with a statistically significant difference. Phospholipids content increased from 6.25 ± 0.7 (µg/g ADW) in NL to 8.23 ± 0.69 (µg/g ADW) in BL. However, this is not significantly different when measured on protein basis according to one-way ANOVA and Duncan Test.

The significant difference in MDA content in NL and BL was highlighted in Figure 5.5. There was a statistically significant difference in MDA content between NL and BL extracts on both ADW and protein basis. MDA content decreased dramatically from 4.7 ± 0.89 (nmol/g ADW) or 2.3 ± 0.53 (nmol/g ADW) to 2.53 ± 0.4 (nmol/mg protein) or 1.09 ± 0.23 (nmol/mg protein).

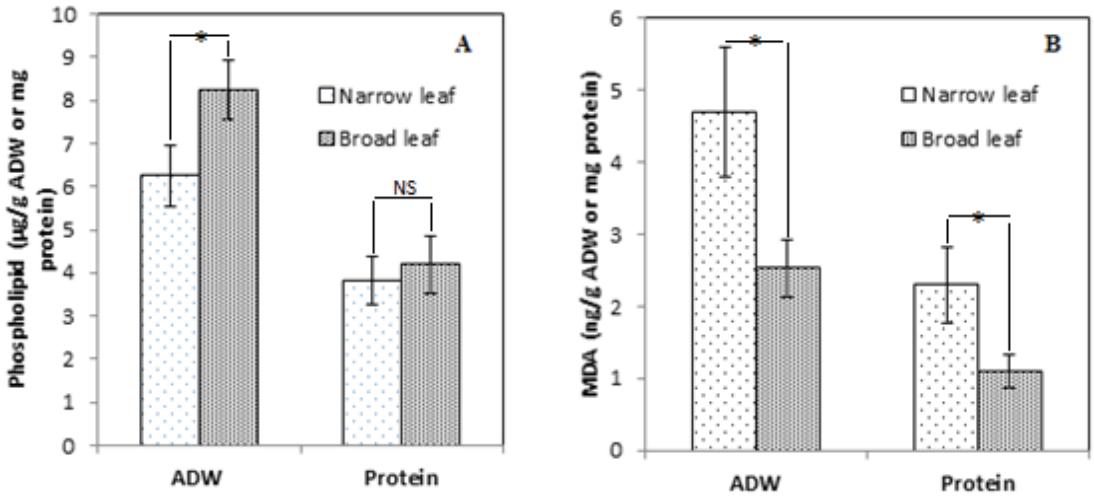


Figure 5.5 Phospholipid (A) and MDA (B) content on a protein basis and a dry weight basis in NL and BL extracts. Mean values ± (n=8). Data and result were based on one-way ANOVA and Duncan Test. An asterisk indicates that treatments were significantly different ($p<0.05$). NS indicates that treatments were not significantly different.

5.4 Discussion

5.4.1 Protein content

Protein content in *C. hindsii* determined by a commercial assay kit (Bio-Rad) showed an elevated level in BL. It is hard to make a direct comparison between protein content of *C. hindsii* leaves and other related plant species using the Bio-Rad reagents protocol to quantitate protein exist because of unavailability of the report in the literature. However, it is higher than soluble protein content in leaves of chives (*Allium schoenoprasum*) cultivated plant which contains less than 2mg/g protein regardless ADW or DW (Stanjner et al. 2011). The higher level of protein in BL in this study may be due to the significant difference in thickness of BL leaves which may contain more amounts of proteins so that plant can perform better in growth and development. Proteins are an integral part of the chemical machinery for cell growth, division, and differentiation in young developing cells. The most important role of the cell proteins is an enzymatic activity which catalyses each reaction step in the biosynthesis pathways that produce the cell components. Other proteins play key roles in signalling transduction that either turn on or turn off the expression of genes in the nucleus by binding to regions of nucleic acid that have a regulatory function (Sjuts et al. 2017; Sjuts et al. 2017). In general, proteins are well-known for their mechanical role which provides a structured framework, or scaffold, for nucleic acids and other cell components (Day 1996; Müntz 2007). Increase in protein levels probably resulted from an increased synthesis and decreased degradation of protein (Kennedy & De Filippis 2004). Therefore, the higher protein content in BL indicated that this phenotype might perform better for physiological functions and further proteomic study should be implemented to understand the mechanism hidden under this variation (Figure 5.2)

5.4.2 Protease

Proteases are rich in leaves of many plant species and conduct several physiological functions, particularly the degradation of proteins. Their activity levels are depended on some internal and external factors, such as the age of tissues, leaf senescence, localisation, and oxidative stresses (van der Hoorn 2008; Simova-Stoilova et al. 2010). In plants, proteases are more often present at a pH range of acid and neutral (Bijina et al. 2011; Maciel et al. 2011), and are less detected at alkaline pH (Gaur & Wadhwa 2008). Also, under oxidative stresses, proteins are often attacked by ROS and can be degraded by plant proteolytic enzymes and provoking dehydration of cells (Kidric et al. 2014). For leaf materials with similar age and not being a state of senescence, it is important to examine how the proteases localisation and other internal and external factors (if any) affect the activity of proteases in a comparing analysis (Figure 5.2).

The activities of proteases were measured in NL and BL phenotypes of *C. hindsii*, and the results showed an around three times stronger activity of acid protease (AP) over neutral proteases (NP) in both phenotypes. This finding was in line with other proteases studies on medicinal plant of drumstick tree *(Moringa oleifera)* and cereals like rice *(Oryza sativa)* in which most of the measurable proteolytic activities of leaves are accounted for by a number of acid endopeptidases (Fukayama, Abe & Uchida 2010; Bijina et al. 2011). Most protease activities occurred in crude plant extracts localised in vacuoles where contain all types of proteases active optimally at acidic optima as vacuole compartment represents a major site of amino acid recycling in response to plant development and environmental effects (Müntz 2007). They probably are either aspartic peptidase (APs) (pepsin-like proteases) and cysteine peptidases (CPs) (papain-like proteases) or even carboxypeptidases (belong to serine peptidases (SPs)) mostly active at acidic pH (van der Hoorn 2008; Rawlings et al. 2014). In contrast, the majority of SPs are active within the neutral to the alkaline region but rarely found in plants. This pattern of protease localisation is relevant to our study results where AP activity was found significantly active and distinguished in NL of *C. hindsii*.

The endogenous activities of AP and NP in the leaves of both NL were distinguished in patterns and magnitude. AP activity was not only higher than NP in both phenotypes but also significantly elevated in NL, while there was no significant difference in the activity of NP (the AP and NP were calculated and consistent between ADW and protein basis). It is expected that there are possible interactions between acidic proteases species formed and released during cell wall degradation and tissue desiccation induced by oxidative stresses occurred in NL. Several recent studies have indicated a major role for the protease in the plant leaves responding to oxidative stresses. Increases in proteolytic activity have been correlated with the liberation of α-NH_2 nitrogen, characteristic of oxidative stresses induced enzymatic processes (Palma et al. 2002). The hypothesis developed by Kidric, Kos & Sabotic (2014) was that intracellular proteins are oxidatively modified by free radicals (or maybe by related oxidants), followed by selectively recognised by intracellular proteolytic enzymes. Acidic protease increased in the wild-type *Arabidopsis*

leaves upon exposure to excessively strong illumination (Kato et al. 2013). Simova-Stoilova et al. (2010) indicated the increasing involvement of acidic proteases in water stress of *Triticum aestivum*. Drought-induced an increase in endoproteolytic activity of the aspartic protease in leaves of *Phaseolus vulgaris* (Roy-Macauley et al. 2001b), and proteases were identified up-regulated as their activity in Rubisco protein degradation in leaves of *Medicago sativa* (Aranjuelo et al. 2011). Rubisco degradation was also found by the increase in cysteine protease activity in leaves of *Ricinus communis* in response to methyl jasmonate stress and mechanical wounding (Maciel et al. 2011).

There has been a greater number of studies identifying the correlation between drought-induced stress which is the most frequently studied abiotic stress and protease activity (Kidric et al. 2014). In a natural environment where factors that induce a state of stress seldom act individually, however, different stresses can have the same effect at the cell level and plant responses to them perhaps with the same molecular mechanisms (Kidric et al. 2014). Furthermore, a combination of abiotic stress induces production of ROS, whose levels can increase substantially causing damages to cell structures, including proteins. ROS brings about inhibition of protein synthesis or causes protein denaturation (Das & Roychoudhury 2014). In an earlier study, Fazeli, Ghorbanli & Niknam (2007) reported that total protein content in leaves of two sesame cultivar (*Sesamum indicum* L.cvs. Darab 14 and Yekta) decreased obviously under drought stress (Fazeli et al. 2007). Drought stress also induced protein loss in leaves of drought-sensitive cultivar of common bean (*Phaseolus vulgaris*) (Roy-Macauley et al. 2001) and leaf of wheat (*Triticum astivum*) (Simova-Stoilova et al. 2010). In the current study, desiccated state of all leaf material is considered as a major source causing ROS which may lead to a decrease in protein levels, especially in NL. Many of these events are known to induce proteolysis, uncontrolled or regulated and limited (Kidric et al. 2014). Although, it is complicated to differentiate between the different plausible causes of change in acid protease activity in narrow leaf and broad leaf cultivars of *C. hindsii* especially when the leaf tissues of both have been experienced through a chronic desiccated state.

Acid and neutral protease activity and protein content were found related to *Grevillea* leaf tissue (Kennedy & De Filippis 2004) where low protein content in leaf tissue was associated with high protease activity. Accordingly, in the case *C. hindsii*, protein content was decreasing in the event of increasing acid protease activity in NL cultivar. On the other hand, responding to a slight decrease of neutral protease activity in narrow leaf, protein content was increasing but not statistically different. The relationship between protease activity and protein content is not solely the component of degradation in the desiccated tissue under combined oxidative stresses. It has been reported that proteases do however play an important role as terminal protein degraders and controlled by intracellular compartmentalisation and some endogenous inhibitors (Kidric et al. 2014). On the other hand, the modification of proteins including protease proteins promoted directly by oxidative stress and can be measured by the production of carbonyl groups in the molecule (Reinheckel et al. 1998; Kingston-Smith & Foyer 2000). Therefore, the number of carbonyl groups generated under stresses may be the ruler to measure

proteolytic activity. However, this is upon the scope of this study; it is worthy to investigate in a future study.

5.4.3 Proline and glutathione

a. Proline

Proline is well-known amino acid because it has several important functions, such as acting as an organic osmoprotectant, chelating metals, inhibiting lipid membrane peroxidation, and is considered as a potent antioxidant (Gill & Tuteja 2010; Szabados & Savouré 2010; Burritt 2012). Therefore, it is important to investigate the concentration of this multi-functional amino acid in a commercial medicinal plant, particularly in different phenotypes of *C. hindsii*. The result is not only important in selecting a superior plant phenotype for breeding and cultivation but also to understand the mechanism of action in which proline responses to, such as oxidative stresses.

In leaves of *C. hindsii*, proline content was significantly higher in NL based on ADW basis but showed a similar concentration calculated on protein basis. Information about proline concentration has been limited in previous *in vivo* studies on *C. hindsii* and related medicinal plant species but varies among other medicinal plants in *in vitro* studies, which investigated the relationship of proline levels with other phytohormones and oxidative stresses. In such contexts, proline accumulation has a complex effect on plant development and is susceptible to other adverse environmental conditions.

Proline can affect both positively and negatively by interaction with other phytohormones under stresses (Iqbal et al. 2014). Gibberellin acids (GA) which regulate major aspects of plant growth and development increased proline content in response to salinity (Tuna et al. 2008), whereas *Zea mays* and *Anabaena* plants treated by GA reduced the proline accumulation also under salinity stress (Tuna et al. 2008; Alia & Gahiza 2007). Abscisic acid (ABS) was directly involved in osmo-induced proline accumulation in *Brassica rapa* leaf discs (Trotel-Aziz et al. 2003) and accumulated in both ABA-treated and untreated maize cells in response to chilling (Chen & Li 2002). However, it was reported that exogenous ABS treatments did not affect proline accumulation in *Spinacia* or *Pennisetum* seedlings according to McDonnell, Coughlan & Wyn-Jones (1983). According to Manjili, Sedghi & Pessarakli (2012), salinity stress decreased the activity and proline content and other antioxidant enzymes of various wheat cultivars. It is suspected that other phytohormones probably regulated proline accumulation variation between NL and BL of *C. hindsii* under either specific oxidative stress or by total effect of adverse environmental conditions. Also, most of the above studies were conducted in *in vitro* condition, which is usually not consistent with *in vivo* studies because of their different levels of controlled abiotic conditions. Therefore, the upregulated level of proline in NL may be contributed by some internal factors, such as leaf characteristics.

b. Glutathione

Glutathione (GSH) is widely used as a biomarker of oxidative stress in plants, although its role in plant metabolism is a multifaceted one (Tausz, Sircelj & Grill 2004; Gill et al. 2013; Wujeska-Klause, Bossinger & Tausz 2015). GSH is one of the most crucial metabolites in plants and plays a key role in the antioxidant defense system against ROS induced oxidative damage. However, the concentration of cellular GHS greatly varies under abiotic stresses and considerably contributes to its antioxidant function. Recent studies have confirmed that the concentration of GSH increases in plant exposed to cadmium (Sobrino-plata et al. 2014), (Mishra et al. 2013), drought (Mekki, Hussien & Salem 2015; Samuilov et al. 2016), and heat stress (Wujeska-Klause et al. 2015). In contrast, increased level of GSH did not enhance resistance to Cd tress, and cold pre-treatment caused a slight decrease in the glutathione pools of *Dunnaliella viridis* (Madadkar et al. 2014); water deficit caused a reduction in GSH in seedlings of rice (*Oryza sativa*) (Pyngrope et al. 2013). In this study, GSH content was similar between NL and BL of *C. hindsii* and this similarity may indicate that GSH might not associate with stresses occurred in two cultivars (if any) or its activity might rely on several factors which are not related to the variations in NL and BL. In comparison with leaves of *Allium schoenoprasum* cultivated plants (Stanjner et al. 2011), the GSH quantities in *C. hindsii* was much lower (with a maximum of 12.97 nmol/mg protein *versus* 11.76 x 10^3 nmol /mg protein in *A. schoenoprasum*).

Leaf materials of both cultivars of *C. hinsdii* were under desiccated and freezing (-80 degree) state for a period, and may result in damages in plant cells and changes in metabolism, leading to oxidative stresses by the formation of radicals and increased solute concentrations (Burritt 2012). Under cryopreservation, proline has demonstrated its important role in plant cold tolerance (Hoffman et al. 2010; Burbulis et al. 2011). Protective role of GSH has been observed against cold hardening by increasing in content in the shoots of wheat (*Triticum aestivum* L.) (Kocsy et al. 2000) and seedlings of (*Jatropha curcas* L.) (Fan 2013); high GSH content was detected in spruce during winter (Polytechnic 1993). In other experiments, Lukatkin, Anjum & Mordovia (2014) revealed that GSH has a high potential for sustainably increasing chilling resistance in plants.

Also, under dehydrating conditions, plant cell often undergo osmotic adjustment by accumulating several small molecular weight organic solutes, and proline is one of the most common and preferred osmolytes in many plants because it does not interfere with normal cellular processes and biochemical reactions (Burritt 2012). In *Triticum aestivum* plants (durum wheat), proline was found positively related to osmotic adjustment under salinity stress (Poustini et al. 2007) and water loss (Ashraf & Foolad 2007). Desiccation induced oxidation of almost all of the GSH, conversion into GSSG occurred (Kranner & Birtic 2005). This result in a 50% loss of the total GSH pool under dehydration development in wheat (*Triticum astivum* L.) however remained higher in tolerant seedlings (Lascano et al. 2001; Gietler 2016). In short, both NL and BL of *C. hindsii* encountered similar conditions under freezing and desiccation leading to the presence of

proline as an effective cryoprotectant while leaf tissues being persevered and GSH as an antioxidant.

5.4.4 Tocopherol and flavonoids

While proline and GHS are popular as biomarkers against oxidative stresses, tocopherol and flavonoids are recognised as strong antioxidants, and have enormous effects on not only plant adaptation and survival but also human health and diets (Xiao, Chen & Cao 2014; Alasalvar & Bolling 2015; Zhang & Tsao 2016; Saini & Keum 2016). Tocopherol and flavonoids contained extracts (only being biosynthesised in plants) are valuable and have attracted scientists and pharmaceutical medicine industry in alternative disease treatments (Mocchegiani et al. 2014; Zhang & Tsao 2016; Cijo, Dellaire & Rupasinghe 2017). Therefore, measuring the concentration of these two plant antioxidants in plant tissues is important in studies on metabolites for sustainable conservation and utilisation.

a. α-tocopherol (Vitamin E)

α-tocopherol as an independent antioxidant or in conjunction with other antioxidants plays an important role in response to stress (Munne-Bosch 2005; Gill & Tuteja 2010; Miret & Munn 2015). This antioxidant deactivates photosynthesis-derived ROS and prevents the propagation of lipid peroxidation as a lipid peroxyl radicals scavenger (Gill & Tuteja 2010). In this study, NL and BL of *C. hindsii* contained significantly different content of α-tocopherol in their leaf tissues (18.81 ± 1.5 vs 22.29 ± 1.23 µg/g ADW, respectively). The maximum content of α-tocopherol was found in the leaves of *C. hindsii* (roughly 0.00223% by ADW or 0.00531% by DW) by colourimetry. Whereas the major industrial source of α-tocopherol is a residue obtained from the distillation of soya bean with only 0.051 – 0.111% (Slover et al. 1983), and in an α-tocopherol-rich plant of *Pistacia lentiscus var. chia* (0.0053%) (Kivcak & Akay 2005), both based on dry weight. Therefore, the leaves of *C. hindsii* may be considered as a potential new source of this vitamin E.

The content of α-tocopherol in BL was measured higher than that in NL based on both means of ADW and protein. Generally, α-tocopherol level changes significantly during plant growth and development, and in response to stressful environmental conditions. However, α-tocopherol does not lead to a significant degradation until the tress is more severe causing an increase of ROS in chloroplasts (Das & Roychoudhury 2014). In other words, when $^1O_2$ is chemically scavenged by α-tocopherol, leading to the formation of α-tocopherol quinone and other oxidation products, which cannot be recycled back to α-tocopherol, and until that time point, α-tocopherol levels will decrease (Munne-Bosch 2005).

The lower level of α-tocopherol level in NL of *C. hinsdii* is perhaps the result of two possibilities: lesser contribution to the protection of α-tocopherol by reducing ROS levels and inhibiting lipid peroxidation; or α-tocopherol degradation exceeds its synthesis and decreases as consequence of severe stresses. That might be explained that the response of

α-tocopherol level is subjected to stress intensity, plant physiological state, and species-specific sensitivity to stress (Munne-Bosch 2005). For instance, high light exposure associated with three times more α-tocopherol content in photosynthetic membranes in leaves of *Fagus sylvatica* than those kept in the shade (Lichtenthaler 1979), and higher content of α-tocopherol also found in *Euglena gracilis* (Tanaka et al. 1999; Lichtenthaler 2007). However, an approximate 40% depletion of α-tocopherol was observed in leaves of *Arabidopsis* when exposed to 1500 mmol m^{-2} s^{-1} photosynthetically active photon flux density (PPFD) compared to the intensity of 1000 PPFD (Michel Havaux et al. 2000). Higher tocopherol content usually found in drought tolerance compared to susceptible ones (Munne-Bosch 2005; Buchner et al. 2017). An increase in α-tocopherol occurred as result of damage to chloroplast membranes by UV-B radiation in several species (Delong & Steffen 1997) or under high photosynthetic irradiance (Lichtenthaler 2007). A study regarding salt tolerance showed that α-tocopherol play an enhancement role and alleviate the harmful effect of salinity stress in flax plant (*Limum usitatissium*) (Sadak et al. 2014). α-tocopherol responded differently in plants species such as increasing in maize (Leipner et al. 1999) but declining in tomato (Walker & McKersie 1993) when both expose to low temperature. In short, it is still uncertain to conclude which mechanism possibilities led to different levels of α-tocopherol in NL and BL of *C. hindsii* as both of the above can be possibly occurred. However, it is important to state that BL possesses a better qualitative trait regarding the content of α-tocopherol.

Although α-tocopherol plays a major role in plant stress tolerance, others studies revealed that of α-tocopherol is not essential for plant survival as its deficiency induces only a slightly increased susceptibility to photo-oxidative stress (Kanwischer et al. 2005). It is predicted that the distinguished levels of α-tocopherol in NL and BL exhibits not only the response of α-tocopherol to stresses but also the relationship with other substances in the antioxidant network, such as ascorbate, glutathione, flavonoids, and terpenoids (Sergi 2005). This hypothesis has been proven by some studies: a mutant of *Arabidopsis* showed a net α-tocopherol loss under severe deficit induced by a deficiency of ascorbate in chloroplasts (Munne´-Bosch & Alegre 2002). In response to the light stress of another mutant of *Arabidopsis* which is deficient in β-carotene, α-tocopherol accumulation increased during acclimation to high light (M. Havaux et al. 2000).

In another study on the effect of drying methods on antioxidant content, α-tocopherol was found to be the most stable vitamin under all drying conditions (86.4 % retention during oven drying) (Saini et al. 2014). The collection and preservation of all leaf tissues were conducted by the same condition on silica gel drying at -80°C storage. Therefore, the α-tocopherol levels were perhaps affected slightly, and its different levels in NL and BL must be due to their natural conditions where they were cultivated.

In conclusion, α-tocopherol is generally related to stresses by changes its level depending on the species tolerant or susceptible to stress. These changes, however, are the consequence of more complicated mechanism, where α-tocopherol protects the plant from abiotic-induced stress and is limited by the degree of stress and the relative amounts

of other antioxidants (Sadak et al. 2014; Hussien, Salem & Mekki 2015). On the other hand, α-tocopherol also modulates other antioxidants and confer better tolerance against stresses (Kumar et al. 2013). Therefore, it is necessary to consider the response of α-tocopherol together with the other antioxidants. As the limitation of this study, apart from proline, glutathione, further assessments will be implemented on other antioxidants such as flavonoids, phospholipid and measuring lipid peroxidation through the activity of MDA.

b. Flavonoids

In plant, flavonoids are widely distributed and have important roles in plant physiology. Flavonoids pathways have been confirmed in various plant species to respond to almost all abiotic stresses such as strong light, ultraviolet (UV) radiation, low/high temperature, ozone, heavy metals, drought, *etc.* (Winkel-Shirley 2002; Mierziak et al. 2014). The current study results showed that total flavonoids accumulation was accelerated in BL of *C. hindsii*. The result indicates that elevated level of flavonoids in BL may associate with the response of plant under abiotic stresses. Regarding defense-related functions, flavonoids can be divided into two groups: "preformed" and "induced" compounds. While the "preformed" flavonoids are synthesised during the normal development of plant tissues, the "induced" ones are synthesis by the plant in response to physical injury, infection, or stress (Treutter 2006). A higher level of flavonoids accumulations was measured in BL compared to NL (23.16 mg/g ADW versus 18.75 mg/g ADW or 40.30 mg/g DW versus 22.10 mg/g DW[1]). Due to BL and NL plants were in *in vivo* condition, the flavonoid differences between BL and NL may be contributed by both "performed" and "induced" sources. The flavonoid content in *C. hindsii* is relatively high compared to other medicinal plants (Table 5.2). However, due to analytical variation (e.g. extract solvents and standard reagents), these measurements can result in a large range of differences in the amount of flavonoids (Banothu et al. 2017).

Table 5.2 Total flavonoid content in some medicinal plants

Species	Total flavonoid content (mg/g dry weight)	References
Physalis minima	90.64	Banothu et al. (2017)
Melliotus officinalis	57.00	Pourmorad et al. (2006)
Adiantum capillus-veneris	78.30	Pourmorad et al. (2006)
Amaranthus acanthochiton	4.10	Jiménez-Aguilar & Grusak (2017)
Senna Mill (4 species)	29.50 – 77.37	De O Maia et al. (2017)
Eichhornia crassipes	0.49	Tyagi & Agarwal (2017)

[1] The content of flavonoid/dry weigth was converted from ADW based on percentage of water calculated from ADW of C. *hindsii* leaves which contained 42.53% water in BL and 15.14% in NL (Appendix 3).

It is suspected that the greater exposure to sunlight is the main factor regulating the higher level of flavonoids accumulated in BL of *C. hindsii*. However, this assumption does not refuse the role of other biotic and abiotic stress which probably occurred and associated with flavonoid accumulation level during the growth and development of the species. This assumption is in agreement with Kanazawa et al. (2012) that the accumulation of flavonoids increased strongly in response to increased UV radiation. In fact, flavonoids are scarcely produced in plants or organs grown in the dark, because the expression of genes encoding for chalcone synthase (CHS) (the main enzyme involved in flavonoid biosynthesis pathway) is strictly dependent on light (Petrussa et al. 2013). Flavonoids generally absorb UV-B (280-320nm – cause the most severe damage to plant with higher energy compare to UV-A and UV-C regions) and protect the tissue from damage as a UV-filter. UV-B radiation can disable the processes of cell division and alters the pattern of plant growth (Jaakola & Hohtola 2010; Agati et al. 2013). As a result, the growth of UV-B radiation sensitive plants might be strongly affected resulting in morphological changes (Amalesh et al. 2011). In contrast, the response of plants to more intense UV-B radiation resulting in a reduction in radiation absorption and severe damages, have reduced level of flavonoids than in those with normal pigment complement (Agati et al. 2013). However, this is likely not relevant to the lower level of flavonoids in NL because in fact, these individuals are planted in more shaded gardens.

Contributing to the significant role of flavonoids in response to stressful conditions, Ma et al. (2014) stated that these compounds have protective roles in plants which are exposed to drought as the increased total flavonoids accumulation after drought treatment in wheat leaves (*Triticum aestivum*). This accelerated level of flavonoids was also found the same in *Reaumuria soongorica* (Liu et al. 2013), in *Cistus clussii* (Hernández et al. 2004). Low temperature leads to a remarkable increase in phenylalamine ammonia-lyase (an important enzyme in flavonoid synthesis pathway) activity in winter oilseed rape (*Brassica napus var. oleifera*) leaves (Stefanowska et al. 2002). This indicates that flavonoid may perform a functional role in the plant under cold acclimation and freezing conditions in which a large amount of water is removed from the cell into intercellular ice crystal (Amalesh et al. 2011). Therefore, the desiccated and freezing storage state of *C. hindsii* tissues may be responsible for the high level of flavonoids.

In short, accelerated flavonoids accumulation is species specific, and is well-known associated with the response of plant under stressful conditions. Relative high and differentiated levels of flavonoid accumulation in *C. hindsii* in this study is suspected due to abiotic stresses, particularly but not limited to desiccated state and freezing preserved condition of plant tissues, and different intensity exposure to sunlight. However, flavonoid synthesis pathway is a spatial-temporal dependent process which relates to various factors (Pollastri & Tattini 2011), it is uncertain to confirm any change in flavonoid accumulation as a result of any single factor in this study. Also, with the diversity of more than 9000 different flavonoids have been discovered in plants (Mierziak et al. 2014), more flavonoids compounds searches regarding structure, classification,

stress responses, functions, as well as the regulatory and biochemical mechanism need to be conducted in future studies.

5.4.5 Phospholipids and malondialdehyde (MDA)

a. Phospholipids

Lipid peroxidation is a natural metabolic process occurring in plant tissues. However, this process is induced by oxidative stresses and can cause damage integrity of the tissues, particularly the cell membrane, where contains a majority of polyunsaturated fatty acid. These damages are followed by disintegration of organelles, oxidation and dysfunction of proteins (Farmer & Mueller 2013). To reduce the damage and maintain the membrane stability and protection, the presence of phospholipids are believed to be important as they can influence lipid oxidation through a number of mechanisms, such as binding pro-oxidative metals and producing anti-oxidative compounds (α-tocopherol) (Gutteridge 1995; Xue, Chen & Li 2007; Cui, Mcclements & Decker 2015). Therefore, understanding the abundance of phospholipids in plant tissues is valuable to select the better selections for conservation and plant utilisation.

In this study, phospholipids concentrations were observed in leaf tissues of *C. hindsii,* and the results revealed a significant difference in BL (8.23 ± 0.69 µg/g ADW) and NL (6.25 ± 0.7 µg/g ADW). Phospholipids were substantially increased in BL indicating a better possibility of this group against lipid peroxidation in response to adverse environmental conditions.

Among different types of phospholipids, phosphatidic acid (PA) and phosphatidylinositol bisphosphate (PIP_2) have been proven as the key signalling phospholipids in response to stresses (Testerink & Munnik 2011). PA and PIP_2 play an important role in the elevated level of phospholipids measured in various plants in response to stresses (Testerink & Munnik 2011). Under a sub-lethal freezing temperature, the level of phosphatidylcholine (PC), phosphatidylethanolamine (PE), and phosphatidylglycerol (PG) decreased but elevated in PA in *Arabidopsis* (Welti et al. 2002*;* Yu et al. 2010). Salt stress dramatically activates several phospholipids responses, such as PA and PIP_2. However, the other phospholipids remained unchanged in rice leaves (Darwish et al. 2009). PA is a type of phospholipids and constitutes a minor portion of membrane lipids in control condition. However, its formation rapidly increases in response to oxidative stresses (Munnik 2001; Testerink & Munnik 2005, 2011). In plants, lipolytic enzymatic activities of phospholipase D (PLD) are responsible for hydrolysing phospholipids (such as PC and PE) to PA. PA can also be synthesised through the combined action of phospholipase C (PLC) and diacylglycerol kinase (DGK). Therefore, the higher level of phospholipids in BL may be the response of plant against possible oxidative stresses. PA could be responsible for the increase; however due to its minority account in the total phospholipids, it is still challenging to identify which specific phospholipids contribute to their variation levels in leaves of *C. hindsii*.

b. MDA

Malondialdehyde (MDA) is the most frequently measured product of lipid peroxidation as a biomarker of stresses (Tsikas 2017). In many plants, a higher concentration of MDA is generally regarded as elevated oxidative stress (Farmer & Mueller 2013). MDA content in NL was measured almost two times higher than that in BL. Our finding indicates there was an elevated level of lipid peroxidation occurred in NL because of possible oxidative induced stresses. In other words, BL functions a greater role in the protection of leaf tissue membrane. This highlights the role of antioxidants including phospholipids, which involve the membrane protection to reduce the production of MDA (Lin et al. 2014).

In this study, the correlation between MDA content and activity of phospholipids was positive since a larger amount of phospholipids measured in BL. That means NL tissues has undergone more stressful conditions, which accelerated the generation of ROS causing lipid membrane peroxidation. Recent studies have reported that in some plants, an elevated level of MDA in plants was responded to abiotic stresses. Slama et al. (2015) confirmed that water deficit stress led to an increase in leaf MDA content in *Sesuvium portulacastrum*. Drought increased MDA content and lipid peroxidation significantly in leaves and roots of *Sesamum indicum* L. (Suzuki et al. 2014). Another series of experiments on the salt sensitive cultivar of *Sesamum indicum* L. reported a higher accumulated content of MDA compared to the other tolerant cultivars (Gehlot et al. 2005). In the current study, comparing MDA content of Nl and BL showed that cell membrane of BL of *C. hindsii* was probably better protected under oxidative stresses by a higher amount of phospholipids.

The measurement of MDA for lipid peroxidation based on TBARS assay has become a common method as its facile and rapid protocol in which MDA reacts with thiobarbituric acid and the absorbance at 532nm is recorded (Tsikas 2017). However, many plants containing interfering compounds such as carbohydrates and phenylpropanoid-type pigments can also contribute the absorbance at 532 nm (Hodges et al. 1999). *C. hindsii* is a medicinal plant which contains a large number of secondary metabolites such as flavonoids (Hu et al. 2014; Hu et al. 2013; Sung et al. 2008; Thuy et al. 2007; Ly et al. 2006). In this study, the upregulated content of flavonoids in NL might result in an overestimation of MDA in the leaf tissue.

5.5 Conclusion

This study showed the presence of non-enzymatic antioxidant compounds, such as proline, glutathione, α-tocopherol, flavonoids, phospholipids. The content of these secondary metabolites (except proline) were higher in BL compared to NL of *C. hindsii*, indicated that BL should be the selected phenotype for conservation and development of medicinal plant cultivation. The current study has also demonstrated that *C. hindsii* is potential medicinal plant source of phytochemicals that could contribute to human health. The phytochemicals constituents of this plant may be responsible for their efficacy in the treatment of several human diseases. More specifically, BL had a very high concentration

of vitamin E and flavonoids compared to other rich sources of phytochemicals. Because this species is easily cultivated and adapted to a wide range of environments, their cultivation as medicinal sources should be encouraged for commercial purposes, especially in rural areas, where the local livelihood mostly rely on natural resources. Further work has to be carried out to isolate, purify and characterise the phytochemicals of this plant responsible for bioactive properties as well as their molecular mechanisms.

CHAPTER 6

PROTEOMICS

6.1 Introduction

Medicinal plants are among the major and important group of crops which provide a rich source of new vitamins and highly diverse specialised metabolites with important pharmacological properties (Lubbe & Verpoorte 2011; Barata et al. 2016). These plants have been used in various traditional medicinal practices for prevention and treatment of disease for a long time (Chen et al. 2016). In modern medicine, these specialised metabolites have been used as sources of direct therapeutic agents and a destination for new synthetic compounds (Gurnani et al. 2014), which are associated with protective roles on human health and related to a reduction in the risk of cancer, cardiovascular, and other chronic diseases (Jiménez-Aguilar & Grusak 2017; Cijo, Dellaire & Rupasinghe 2017). Also, metabolites are the main means for the plant to adapt to changing and adverse environmental conditions (Mierziak, Kostyn & Kulma 2014; Martínez-Esteso et al. 2015). The most valuable secondary metabolites such as alkaloids, anthocyanins, and flavonoids, have found commercial applications as drug, fragrance, and antioxidants (Jacobs, Heijden & Verpoorte 2000; Mierziak, Kostyn & Kulma 2014). Recently, the demand for large-scale production of these metabolites has been increasing dramatically. Despite the global advanced biotechnology programs and enzyme engineering have achieved, the widespread applications of the engineered plants with improved target metabolites to the industrial implementation remains limited; mainly due to the methodological constraints, such as insufficient genome and metabolic data of plant species and technical failures (Verpoorte et al. 2002; Zhang et al. 2014). Also, most metabolites are characterised by complex chemical structures, regulated by biosynthetic mechanisms involved in complex pathways (Moore et al. 2014). Therefore, better quantitative information and also new knowledge on the cellular mechanisms of such valuable secondary metabolites biosynthetic pathways can improve management of medicinal plant conservation and utilisation.

C. hindsii Benth is an evergreen twining shrub that has been used for generations in Northern Vietnam, particularly by Muong people for diseased treatment relating to ulcers, tumours, and inflammation. Over past decades, *C. hindsii* has been the focus of investigations into a variety of chemical components including sesquiterpene, triterpenes, alkaloids, and flavonoids as well as their bioactivities (Hu et al. 2014; Thuy, Cuong & Sung 2007; Yao-Haur Kuo, Chen & Kuo 1995). Other antioxidant compounds such as vitamin C, vitamin E and some phenolic compounds, which have been shown to have some health effects to reduce the incidence of cancer and cardiovascular diseases, were also isolated from dried leaves of *C. hindsii* (Ly et al. 2006). The rapid increase in the global population with cardiovascular disorder or cancers leads to higher uptake of these medicinal compounds. With a relatively high content of diversely specialised metabolites, the medicinal plant *C. hindsii* is attracting more and more attention internationally.

Recent phytochemical investigations conducted on *C. hindsii* were based on plant materials collected from differently separated locations. Therefore, the results have minimal comparative implications for selection of appropriate varieties for conservation and medicinal plant crops improvement. The occurrence of two different leaf morphological phenotypes (narrow leaf - NL and broadleaf – BL, see Appendix 3) grown in both wild and cultivated conditions may be the result of diverse factors including environmental conditions and molecular regulation. Metabolite investigations in this study have shown that there are significant differences in quantity and activity of vitamin E, flavonoids, phospholipids, proline, and glutathione between NL and BL varieties. Also, protein content in BL was also measured significantly higher than that in NL. However, mechanisms leading to these changes have remained unclear and have received little attention in the literature.

Plants response to changing environmental conditions by reprogramming their cellular machinery at the gene, protein and metabolites levels (Amudha et al. 2005). In other words, the accumulation of secondary metabolites in the medicinal plants is induced by external factors and regulated by complicated molecular mechanisms. To discover these underlying molecular mechanisms, it is important to understand the response of molecules which may regulate the variation of secondary metabolites. The studies of genetic diversity and metabolites (detailed in Chapters 4 and 5 – Enzymes and Metabolites and Genetic Diversity) have provided convincing evidence that high DNA polymorphisms were detected (Appendix 2), and significant difference amounts of metabolites were found between NL and NL. However, from the point of translational to protein levels, it is yet to be ascertained whether those differences are involved certain biosynthesis pathway or is just a consequence of a conditional bias.

Proteomics is the science of proteins, particularly their structures and functions, and is a link between the classical physiological and biochemical methods and nucleic acid tools (De Filippis & Magel 2012). In plants, proteins respond dynamically to biotic and abiotic conditions and are responsible for the functioning and regulation of all biological processes, serve as the basis of the tight homeostasis that characterises any biological system. Therefore, profiling a proteome of an organism can provide insights into the cellular pathways of that species, particularly the biosynthetic pathway of specialised secondary metabolites. In the light of recent plant genomic studies, the power of proteomics by differential displays of the proteome has allowed the identification of polypeptides by mass spectrometer (MS) (Vadivel 2015; Martínez-Esteso et al. 2015). In other words, the dominant approaches for MS analysis heavily rely on the availability of completed annotated proteomes, yet most organisms have no sequenced or poorly annotated genomes and proteomes. Hence it remains challenging to identify proteins and polypeptides from environmental organisms such as medicinal plants (Champagne & Boutry 2013; Jorrín-Novo et al. 2015; Na, Payne & Bandeira 2016).

To overcome the above limitations, several alternative methods have been developed such as *de novo* sequencing, peptide mass fingerprinting (PMF), database searching, and

automation and robotics. *De novo* sequencing is widely used for mapping amino acids from a peptide and inherently requires extensive computation time (Mackey et al. 2002). PMF is suitable to identify protein orientation from organisms whose genomes have been fully described (De Filippis 2013), and automation and robotic procedures are lacking in both experimental procedures and software (Kaczmarek et al. 2002). Protein identification is primarily based on cross-species protein sequence and is helpful tool using database searching strategy (Liska & Shevchenko 2003). Homologous proteins in different species are rarely identical, and the number of sequence mismatches is likely to increase when the phylogenetic distance between species increases. However, a positive protein identification can be obtained if only a few subsets of conserved peptides are identified and conserved (Champagne & Boutry 2013). Recently, the increasing improvements in mass spectrometers have allowed increasing numbers of identified peptides per protein and thus have contributed to more efficient protein identification in non-sequenced plant species (Liska & Shevchenko 2003; Wright et al. 2010). Although wrongly identified protein can occur due to increasing mismatches of identified peptides, the use homologous protein in related species and sequence similarity searches have been widely accepted in a significant number of non-model species. One hundred percent of 80 proteins randomly selected from *Nicotiana tabacum* suspension cells were identified using a database from phylogenetically more distant species, compared to only 11% and 31% of these 80 proteins identified using *N. tabacum*-specific database and *Solanaceae* family database (Witters et al. 2003). Further evidence also showed a higher protein sequence coverage is required for a genetic distance species to acquire a more confident protein identification (Champagne et al. 2012). The major and most comprehensive sequence databases resources have been developed by the National Centre for Biotechnology Information (NCBI) and the Universal Protein Resources (UniProt). While the NCBI database contains a very large number of nucleotide, EST, and protein sequences, UniProt consists of two databases, TrEMBL which contains unrevised, automatically annotated sequences, and Swiss-Prot which contains revised and manually annotated data (Champagne & Boutry 2013). Although Swiss-Prot contains a lower number of entries and is likely unable to keep pace with the increasing number of the sequenced genomes, this database is considered as more reliable resources (De Filippis & Magel 2012; De Filippis 2013). The SIB Swiss Institute of Bioinformatics (SIB) has developed the Expert Protein Analysis System proteomics server (ExPASy; http://www.expasy.org), which is the entry to UniProt. ExPASy tools are designed to deal with several aspects of protein analysis, including BLAST search, proteomics, and sequence analysis, and take into account all splice variants as annotated in UniProtKB (Boutet et al. 2016).

Up to date, there has not been any proteome report from medicinal plant *C. hindsii.* Therefore, the underlying cellular mechanisms leading to a low and high level of several significant metabolites in NL and BL of *C. hindsii* remain elusive at the proteome level. Comparative proteomic analyses have been performed on many crops and medicinal plants to examined differentially expressed proteins under various conditions and in

different plant cultivars and species (De Filippis & Magel 2012; Yang et al. 2015 Wang et al. 2016; Sun et al. 2016; Bryant, Patole & Cramer 2016; Bao et al. 2017; Shi et al. 2017; Naru et al. 2017). However, the proteome of *C. hindsii* responsible for the two phenotypes has not yet been investigated, and such research could be useful to gain a new understanding on the pharmaceutical properties of *C. hindsii* – a potential candidate for medicinal plant crop.

Therefore, the current study, for the very first time to the best of our knowledge, investigates further molecular changes in the plant *C. hindsii* to reveal the protein profile in NL and BL types by adopting a comparative analysis using 2D-PAGE coupled with MS to identify the differentially abundant proteins involved in the two distinguished morphological forms. Combining proteomics with molecular approaches allows a better understanding of the biosynthesis pathway of important metabolites (e.g. flavonoids) and polypeptides of medicinal values.

6.2 Material and method

6.2.1 Plant material

Plant materials of narrow leaf (NL) phenotypes were collected in home gardens in Phu Tho Province. While broad leaf (BL) plants were collected from four provinces (Phu Tho, Ha Noi, Hoa Binh and Hai Phong) in both home gardens and the wild. Leaves were manually harvested from mature plants, cleaned and immediately protected away from moisture and direct sunlight in silica filled sealed plastic bags in the field, before shipment. Air-dried leaves after transporting from Vietnam to Australia with quarantine permission were store in -80^0C freezer in UTS Science laboratory before use. Air-dried leaf material was weighed using an electronic balance (Sartorius, Quintix224-1S, Germany, accuracy ± 0.0001 g), sliced into thin pieces (1 - 2mm^2), and frozen immediately in liquid nitrogen. The frozen tissues were subsequently ground into fine powder by pulverising using a cryomill (Retsch MM200) with a 1cm stainless steel ball. The final powder was then stored in liquid nitrogen for immediate protein extraction or stored at -80^0C for later extraction.

6.2.2 Chemicals

Acrylamide, Polyvinylpolypyrrolidone (PVPP), standard protein molecular weights and carrier ampholyte were purchased from Sigma (St. Louis, MO, USA). SDS, TEMED, ammonium acetate and β-mercapto-ethanol were from Sigma-Aldrich. The immobilised pH gradient, chemical agents, and equipment needed for 2-DE were supplied by Bio-Rad. Double-distilled water was used to prepare all solutions.

6.2.3 Protein extraction

The borax-PVPP-Phenol method (BPP) was proven to be very effective to deal with recalcitrant plant tissues and facilitate downstream applications by removing most of the

interfering compounds and produce high-quality protein samples (Wang et al. 2007). Interfering compounds such as polysaccharides, polyquinones, and phenolic compounds were removed by borax and PVPP. β-mercaptoethanol and ascorbic acid were also added to prevent the oxidation of polyphenols. In this study, the BPP protocol was selected with some modifications; the protein extraction buffer contained the addition of SDS and protease inhibitor, and the protein precipitation reagents (ammonium sulphate saturated-methanol) were replaced by 0.1M ammonium acetate saturated-methanol. These alterations facilitate protein solubility (Yao et al. 2006) and ensure a minimum activity of protein degradation (Wang et al. 2003; Wu, Gong & Wang 2014).

In short, 1g of frozen dry tissue powders were resuspended in 3 ml ice-cold extraction buffer of 100 mM Tris (pH 8.0) containing 100 mM EDTA, 50 mM borax, 50 mM vitamin C, 1% PVPP w/v, 1.5% Triton X-100 v/v, 20% SDS, protease inhibitor, 2% β-mercaptoethanol v/v and 30% sucrose w/v. After the sample was vortexed for 5 min at room temperature, two volumes of Tris-saturated phenol (pH 8.0) were added and then the mixture was further vortexed for 10 min. After centrifugation (4^0C, 15 min, 15,000 x g), the upper phase was transferred to a new centrifuge tube. An equal volume of extraction buffer was added into the new tube; the mixture was then vortexed for 10 min, followed by centrifugation (at the same condition). The upper phase was then transferred to a new centrifuge tube. Proteins were precipitated by adding five volumes of 0.1M ammonium acetate saturated-methanol and incubated at -20^0C overnight or at least 6 h. After centrifugation as described above, the protein pellet was re-suspended and rinsed with ice-cold methanol followed by ice-cold acetone twice and spun down at 15,000 x g for 5 min at 4^0C after each washing, and then the mixture was carefully decanted. Finally, the washed pellet was air-dried, then recovered with lysis buffer UTC 7 (7 M Urea, 2M Thiourea, 0.5% C7BzO) followed by the reduction and alkylation of disulphide bonds in a single step, for 90 min at room temperature, using the reducing agent tributylphosphine (TBP, 5mM) and an alkylating acrylamide monomer (AM, 20mM). The reaction was quenched using dithiothreitol (DTT, 20mM).

6.2.4 Quantification of proteins

Proteins concentration was determined using 1D PAGE and densitometry. This method has a number of advantages: (a) 1D PAGE is compatible with the majority of reagents used to prepare the samples, such as urea, surfactants and reducing agents; (b) the fixing of the gel after electrophoresis removes contaminating substances that would interfere with the other assays; (c) the results of 1D or 2D PAGE are directly comparable, and relative quantitation can be performed; samples can be excised, in-gel digested and subjected to LC/MS/MS. Bovine serum albumin was used as the standard. Soluble protein contents were the results of three separate experiments with three replicates in each (n = 9) and standard errors (SE) of the means included.

6.2.5 Protein fractionation and detection

a. Isoelectric focusing (IEF) and Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE)

2D-PAGE was carried out according to De Filippis & Magel (2012) with some modification. Proteins (300 µg) were analysed using IEF in the first dimension, which based substantially on their charge and pH equilibrium. Immobilised pH gradient (IPG) strips (Bio-Rad, pH 3-10, 11cm) were passively rehydrated with rehydration solution (UTC7) for a minimum of 6 h at room temperature. IPG strips with a narrow linear range of pH (pH 4-7) were used eventually to provide enhanced resolution and more precise isoelectric point (pI) values for protein spots. Isoelectric focusing was conducted at 20^0C with a Protean IEF device (Bio-Rad). The gel strips were subsequently equilibrated in equilibration buffer (6M urea, 250mM Tris, and 2%SDS). IEF conditions were: 100 V - 3000 V (slow ramp for 5 h), 3000 V – 10000 V (linear ramp for 3 h), 10000 V (constant for 10 h). Separation in the second dimension was performed in a vertical Criterion^{TM} precast polyacrylamide gel (Bio-Rad) in MES SDS running buffer (Invitrogen, Life Technologies) in a Midi format electrophoresis systems (Bio-Rad) using following voltage steps: 150 V (0.25 h) and 250 V (maximum 0.5 h) or until the bromophenol blue dye front is at the bottom of the gel. Gels were then fixed with 40% methanol and 10% acetic acid for 30 min before protein staining with Coomassie Staining G-250. Gels were then de-stained, an image obtained using a fluorescence scanner (Typhoon FLA-3500).

b. Protein detection

The gel was calibrated in the vertical sodium dodecyl sulphate (SDS) direction using a wide-range low molecular weight protein marker. The pI of each protein spot was also determined from the IPG strip linear range. To avoid false positives, selected spots that were present or absent in samples examined had to be present in a minimum of three separate 2D gels out of five runs. A total of six 2-DE gels originating from three individual replicates of each phenotype group (narrow leaf - NL and broadleaf - BL) were analysed. The term up- and down-regulated (UR, DR) were also used to describe differentially regulated proteins in NL samples compared to BL samples.

6.2.6 Protein identification and bioinformatics

a. Trypsin In-Gel digestion

Selected spots were manually excised from the 2D gels, digested with trypsin (Trypsin gold-MS grade Promega, Mannheim, Germany), and analysed by LC/MS/MS according to Pokharel et al. (2016). Briefly, the gel was sectioned into 1 x 1 mm pieces which were de-stained by adding 50% acetonitrile (ACN)/50mM NH_4HCO_3 and incubated for 10 min at ambient temperature. The process was repeated until the stain disappeared. 100% ACN was added to dehydrate the gel pieces before rehydration with 12.5 ng/µl of trypsin solution and overnight incubation at 37^0C. After incubation, the supernatant was

collected and subjected to 10 min sonication in a water bath, followed by another sonication, then 30 µl of 50% CAN/2% formic acid was added. The solution was added to the previously collected peptides samples, and the volume reduced to 15 µl by rotary evaporation. The peptide solution was centrifuged at 14,000 x g for 10 min to remove any insoluble material before LC/MS/MS analysis.

b. LC/MS/MS

Using an auto-sampler, connected to a nanoLC system (Tempo Eksigent, USA), 10 µl of the sample was loaded at 20 µl/min with MS loading solvent (2% Acetonitrile + 0.2% Trifluoroacetic Acid) onto a C8 trap column (CapTrap, Michrom Biosciences, USA). After washing the trap for three min, the peptides were washed off the trap at 300 nl/min onto a PicoFrit column (75 µm × 100mm) packed with Magic C18AQ resin (Michrom Biosciences, USA). Peptides were eluted from the column and into the source of a QSTAR Elite hybrid quadrupole-time-of-flight mass spectrometer (Applied Biosystems/MDS Sciex) using the following program: 5–50% MS solvent B (98% Acetonitrile + 0.2% formic acid) over 8 min, 50–80%MS buffer B over 5min, 80% MS buffer B for 2min, 80–5% for 3min. MS solvent A consisted of 2% acetonitrile + 0.2% formic acid. The eluting peptides were ionised with a 75 µm ID emitter tip that tapered to 15 µm (New Objective) at 2300V. An Intelligent Data Acquisition (IDA) experiment was performed, with a mass range of 375–1500 Da continuously scanned for peptides of charge state 2+ – 5+ with an intensity of more than 30 counts/s. Selected peptides were fragmented and the production fragment masses measured over a mass range of 100–1500 Da. The mass of the precursor peptide was then excluded for 15 s.

c. Protein identification

Peptides were identified, and protein identity inferred using both MS and PEAKS Studio software (Peak Studio 7.5, Bioinformatics Solution Inc., Waterloo, ON, Canada). The MS and MS/MS data produced by the QSTAR were searched with a taxonomic restriction of *Viridiplantae* (green plants) using the software Mascot Daemon (version 2.4) and searched against the MSPnr100 databases (comprised of all known reference protein sequences including NCBI, RefSeq, UniProt, EuPathDB and Ensembl, containing 75,925,788 sequences; 27,045,014,025 residues). The settings used were as follows - Fixed Modifications: none; Variable Modifications: carbamidomethyl, propionamide, oxidised methionine; Enzyme: semi-trypsin; Number of Allowed Missed Cleavages: 3; Peptide Mass Tolerance: 100 ppm; MS/MS Mass Tolerance: 0.2 Da; Charge State: 2+and 3+.

d. Bioinformatics analysis

The results of the search were then filtered by including only protein hits with at least one unique peptide (Bold Red) and excluding peptide hits with a $p > 0.05$. The Mascot score with at least 15 was considered as a match of proteins. Peptides identified by Mascot were further validated by determining similarity with known proteins in the NCBI database

using protein BLAST based on E-value generated and % similarity (Liska & Shevchenko 2003; Habermann et al. 2004; Buts et al. 2014; Armengaud et al. 2014; Upendar 2017). The threshold was set to a minimal significant of 1e-3 and an identity percentage of >75%. Further protein identification were performed by means of mass fingerprint data, pI, and molecular weight with the MultiIdent tool in ExPASy (www.expasy.org/tools/(Wilkins & Williams 1997; Wilkins et al. 1998; Witters et al. 2003; De Filippis & Magel 2012) to search for close matching protein across species. The theoretical pI and molecular weight of the BLAST hit was calculated using ExPASy tool (http://web.expasy.org/compute_pi/). Cross-species comparisons between organisms that have fully sequenced genomes suggest that amino acid composition and molecular mass are generally well conserved in homologous proteins across phylogenetic boundaries, presumably because of shared domains, and this provides a basis for such comparative proteomic experiments.

To determine the biological role and function, the identified proteins were assigned to Gene Ontology (GO) using Blast2GO software (https://www.blast2go.com/), coupled with the UniProt GO annotation program. The GO database, BLAST annotations, and information reported in the literature were used to analyse and classify the identified proteins based on their cellular localisation, molecular functions, and biological processes. Based on the searches, specific up-regulated and down-regulated proteins were identified.

6.3 Results

6.3.1 Protein extraction

The leaf tissues of medicinal plants contain a relatively low quantity of proteins but a high concentration of protease and other compounds such as phenolics, polysaccharides, lipids, and nucleic acid. Proteins/enzymes involved in the biosynthesis pathway of secondary metabolites are therefore affected by several biotic and abiotic factors. The method of choice for protein extraction in secondary metabolite biosynthesis studies must provide a broad spectrum of protein abundance and successfully remove interfering compounds. Borax/PVPP/Phe extraction which integrated the advantages of phenol-based protocols and TCA/Acetone methods has showed to be the method of choice as it was effective to deal with recalcitrant leaf tissues (Wang et al. 2007; Wang et al. 2013; Zhang et al. 2014; Li et al. 2015). The method was conducted with air-dry weight (ADW) leaf tissues of the medicinal plant *C. hindsii* with some modifications and provided high protein yield and good quality.

Protein yields from BPP of leaf tissues of four species including *C. hindsii* extracted from this experimental study and another three plant species by Wang et al. (2007) are listed in Table 6.1. The average protein yield in *C. hindsii* was approximately 2.35 mg/g fresh weight which is consistent with the previous report by Wang et al. (2007), although there was a slightly higher than the other plants. This higher yield is probably due to the modification of the method which was applied to this medicinal plant with a rich source

of secondary metabolites; such as the addition of SDS and protease inhibitors. In addition, the other plant species, particularly woody and recalcitrant plants, may contain larger amounts of polysaccharides, polyphenols, which may obstruct protein isolation and produce less yield (Alam et al. 2013).

Table 6.1 Protein yield from different plant species extracted by BPP (Wang et al. 2007)

Plant species	Tissue	Protein yield (µg/g FW)	Note
Celastrus hindsii	Leaf, fully expanded	2347 ± 484	Medicinal plant Our extraction (a)
Arabidopsis thaliana	Leaf, fully expanded	2250 ± 321	Model plant
Prunus persica	Leaf, young	2176 ± 288	Woody plant
Pinus bungeana	Needle, mature	1960 ± 352	Recalcitrant plant

a) *Protein yield of C. hindsii was converted from µg/g ADW to µg/g FW as 22,3% water was lost during the drying period.*

Using the BPP method, proteins extracted from *C. hindsii* leaves were distinct between medium and lower molecular weight (M_t) regions of the gel (Figure 6.1). However, bands at the higher molecular weight than 40 kDa were unlikely obtained. These results were consistent with previous findings indicating that some large proteins were lost while many of smaller proteins were enriched (Carpentier et al. 2005).

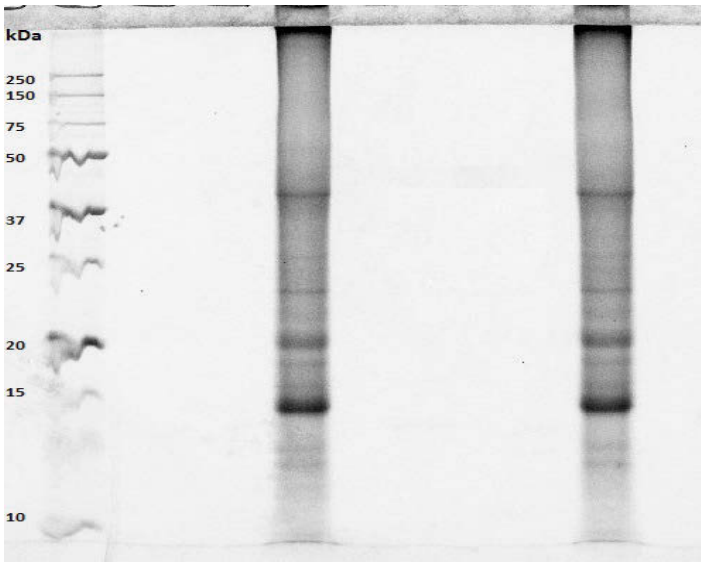


Figure 6.1 Protein profile: 1 D gel with two replicates showed the abundance of low molecular mass proteins but lack of high molecular ones

6.3.2. Proteomic profiling

To understand the possible molecular mechanism of different leaf phenotype forms and variation in secondary metabolites levels in *C. hindsii*, proteome profiles were compared between NL and BL. The 2D gels of proteins extracted from *C. hindsii* leaves collected from NL and BL are shown in Figure 6.2.

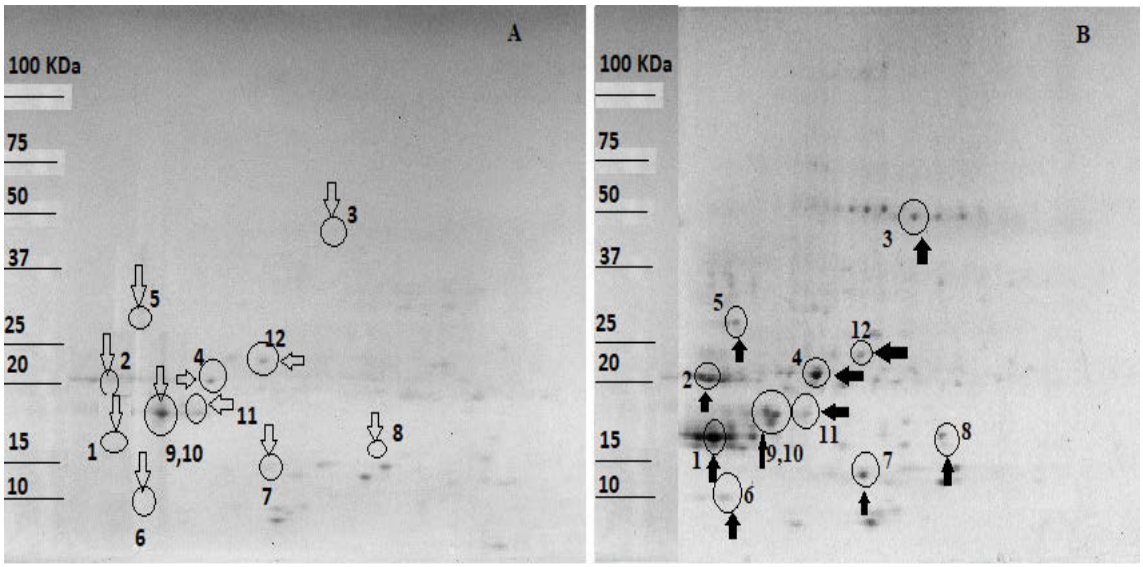


Figure 6.2 Protein distribution and profiling of *C. hindsii* expressed in NL (Figure A) and BL (Figure B) phenotypes. (upper pointed arrow: up-regulated; levelled arrow: present in both NL and BL; down-pointed arrow: down-regulated)

At the proteome level, the protein pattern indicated that a significant difference was observed in NL and BL. The much less protein abundance in NL reflected the condition of preserved material which was severely experiencing a domesticated state, especially for thinner leaves of NL. Consequently, protein degradation probably occurred more severely in NL. Due to experimental variations caused by instrument uncertainty in operation, as often experienced not all protein spots were detected on each gel of the same sample. To average out this experimental variation, five gels were made for each sample. Only spots presented in at least three gels were chosen. Comparative analysis of the proteome visualised in 2-DE gels (pI 4-7) revealed that a total of 8 spots underwent changes in volume variation (1.5-fold, $p < 0.05$) in BL compared to the NL; 2 spots newly appeared in BL, and other 2 significant spots which were expressed with high density in the 2D gels of both NL and BL were also selected for identification. All selected spots were successfully characterised by LC-MS/MS and identified by the software Mascot to search proteins with a taxonomic restriction of *Viridiplantae* (green plants) against the MSPnr100 databases. Due to the extremely poor genome and protein sequence information currently available on *C. hindsii* and the family Celastraceae, identification relied almost exclusively on homology to proteins identified from other plant species. Therefore, a comprehensive strategy for database searching was developed as follows: (a) If a significant score was obtained for one or more peptides, error

tolerant search was performed on the sequence of the identified protein in order to look for additional homologous peptides containing a modification, e.g., an amino acid substitution; (b) if MS/MS ion search against MSPnr100 failed, data were re-searched against all Viridiplantae entries in the UniProt database via the web server ExPASy at (www.expasy.org/tools/) to calibrate peptide mass fingerprinting data; (c) when all database-related strategies failed, fragmented peptides were used to perform homology search in the non-redundant database from NBCI with the MSBLAST search engine.

By using the database search with the method above, 12 protein spots could have possible identification (Table 6.2). By initial search against MSPnr100 databases, proteins were identified from spot 2,3,4, 9, 10 and 11. The identification of the remaining spots failed in 6 cases because of poor MS data. Sequentially, the efficacy of MS/MS fingerprinting coupled with AA composition, pI, and Mr information for identification was conducted on all 12 spots as an alternative strategy and as additional confirmation. Because proteins of interest in *C. hindsii* were likely not to be present in the database, protein identification by peptide mass fingerprinting oftenn requires highly homologous proteins from other species whose proteomes are available in the databases. Furthermore, as demonstrated by Jacobs et al. (2005) even high sequence homology to a database entry might not be sufficient to guarantee identification due to homology not being conserved at the tryptic digest level.

The cut-off of MS/MS scores was 60 and that of BLAST score was 500. For proteins met both criteria, was classified as high scored, and adversely as low scored. For those scored above only the MS/MS or BLAST cut-off were categorised as intermediate group (Table 6.3). To compare between molecular weight and pI values expressed in 2D-PAGE gels and the ones in databases, these two parameters were searched against the Swiss-Prot and TrEMBL (*Italics*), and MSPnr100 databases with limitation on the taxonomy of embryophyta. The results showed the matches of proteins 2, 3, 4, 9, 10, and 11 as these two values are within the range of the available data (Table 6.3). On the other hand, several proteins belonging to intermediate and low score groups are likely not matching as high pI and molecular differences. However, these differences are based on theoretical values and very few embryophyte organisms, therefore future experiments need to be carried out for confirmation of identification.

Table 6.2 Identification of differentially expressed proteins between NL and BL. The cut-off for query cover is 82% and identify is 75%, however, most of proteins have 100% of these two values. UR: Up-regulated; PB: presence on both NL and BL.

Spot no.	Score MS/MS	Protein name	Regula-tion	Score BLAST	E-value
1	40	Inorganic phosphate transporter 1-7	UR	45.2	1e-05
	35	IQ-DOMAIN 1-like		40.5	3e-04
	31	F-box/kelch-repeat protein At5g51250-like		40.9	9e-08
	30	Putative nuclease HARBI1		48.1	7e-07
2	99	Acidic endochitinase	UR	560	0.0
3	823	ATP synthase CF1 alpha subunit	UR	827	0.0
4	86	Chlorophyll a-b binding protein of LHCII	PB	839	0.0
5	29	F-box/FBD/LRR-repeat protein	UR	53.2	3e-07
	17	Hypothetical protein AXG93_1154s1520		50.7	2e-06
	15	Beta-1,3-galactosyltransferase 14		54.5	2e-07
	15	Transportin MOS14		49.8	5e-06
6	24	Ycf1 (chloroplast)	UR	91.8	5e-20
	22	Uncharacterised protein At1g04910		45.6	8e-05
	20	Retrovirus-related Pol polyprotein		57.9	1e-08
7	23	Uncharacterised protein LOC105059940	UR	100	6e-22
	20	Myosin-2-like isoform X1		60.0	2e-09
8	26	F-box protein At1g70590	UR	53.7	3e-07
	24	Endoribonuclease Dicer homolog 3a		49.4	5e-06
	22	Staphylococcal nuclease domain protein 1		50.3	4e-06
9	96	Kunitz family of proteinase inhibitor	UR	389	7e-137
	64	S locus-linked F box protein		723	0.0
	63	Pentatricopeptide repeat-containing protein		985	0.0
	58	Germin-like protein 11		360	9e-125
	55	Cullin-4		1509	0.0
	53	Leucoanthocyanidin reductase 1		662	0.0
10	66	Pentatricopeptide repeat-containing protein	UR	985	0.0
	66	S locus-linked F box protein		749	0.0
	60	Germin-like protein 11		360	9e-125
	49	Disease resistance protein RPS4-like		1488	0.0
	49	Cullin-4		1509	0.0
	48	Hypothetical protein VOLCADRAFT		2001	0.0
	33	Elongation factor G family protein		1302	0.0
11	108	Germin-like protein 11	PB	360	9e-125
	61	Pentatricopeptide repeat-containing protein		985	
	54	ATP synthase subunit d, mitochondrial		985	0.0
	54	Auxin-binding protein ABP19a		341	1e-118
	51	Autophagy-related protein 8D-like		412	2e-145
	51	Uncharacterised protein LOC106400730		268	7e-91
	51	Probable disease resistance protein		866	0.0
	46	Somatic embryogenesis receptor kinase 1		597	0.0
	27	Uncharacterised protein LOC105157445		450	4e-159
12	33	Protein trichome birefringence-like 10	PB	1483	0.0

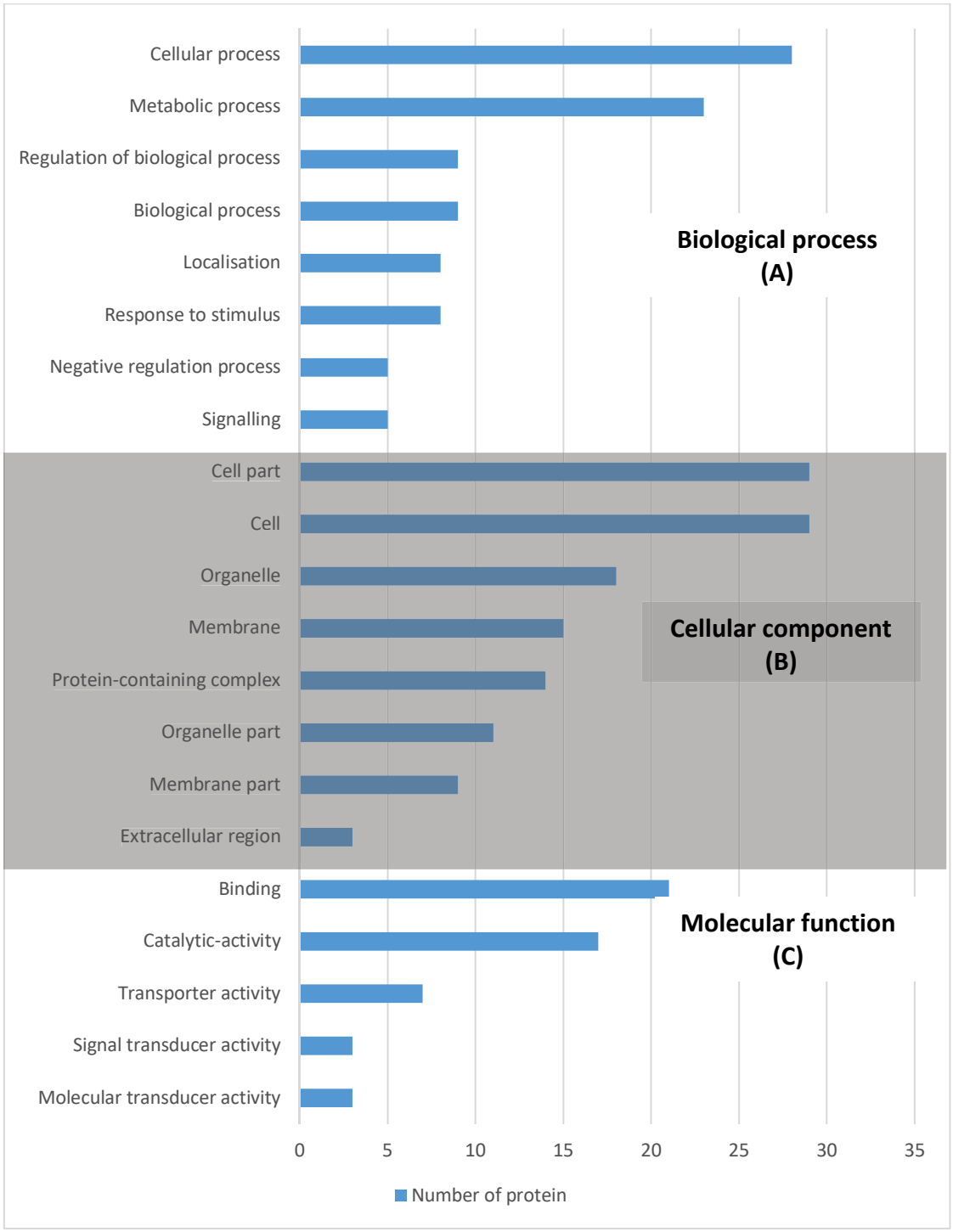


Figure 6.3 Functional classifications of differentially expressed proteins from NL and BL of *C. hindsii*. Proteins were categorised according to their gene ontology (GO) using Blast2Go software for their Biological process (A), cellular component (B), and molecular process (C).

Despite, the MS/MS fingerprinting method from this study were successful in several cases. The use of mass fingerprinting peptide data using the MultiIdent tool in ExPASy

at (www.expasy.org/tools/) were successful in identifying proteins ATP synthase CF1 alpha subunit and Chlorophyll a-b binding protein of LHCII type 1 (corresponding to protein single spot 3 and 4, respectively), germin-like protein, ATP synthase subunit d, autophagy-related protein (in large spot 9, 10, 11, respectively). However , this method failed to generate close matchings of other proteins as indicated by the parameters of score and rank. According to Wilkins & Williams (1997), the confident protein identification should be obtained from the first rank matchings whose scores are less than 30 and their ranks are far different the sequential matchings.

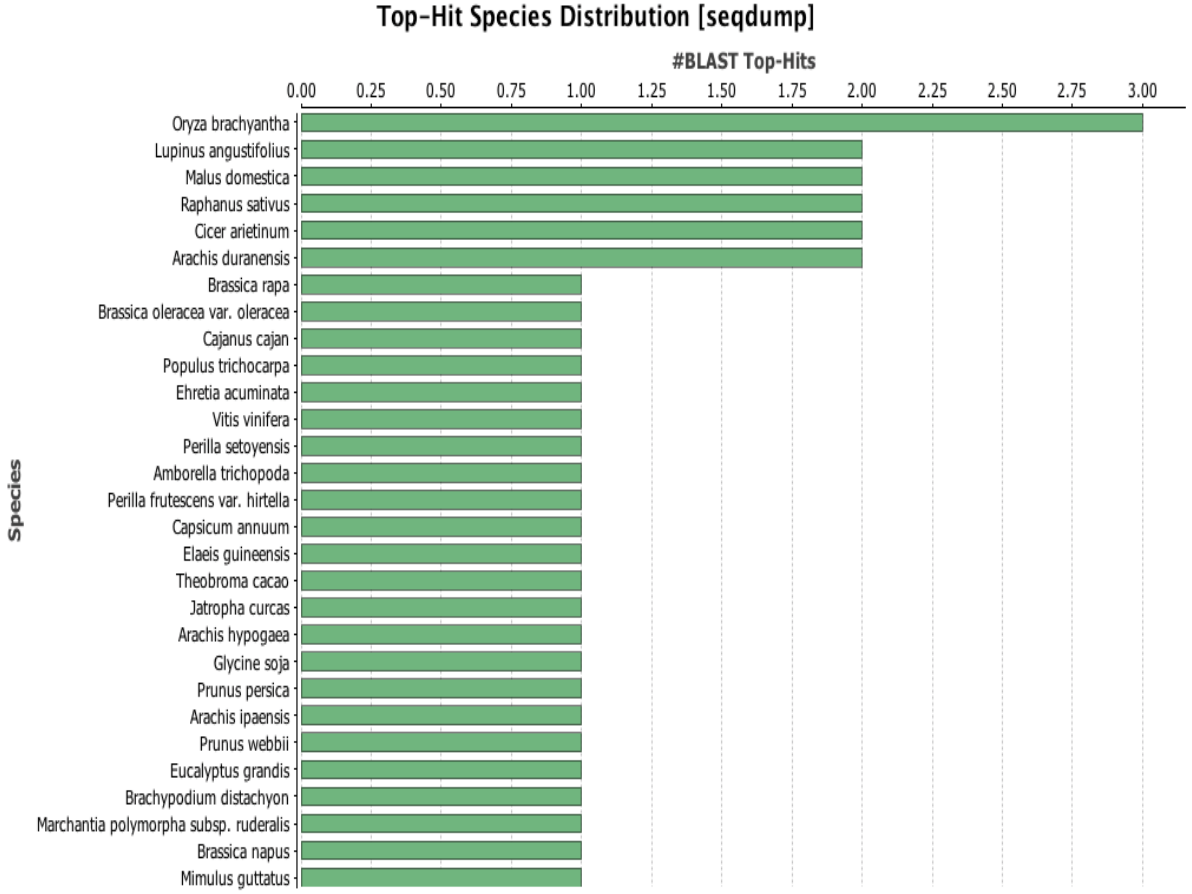


Figure 6.4 Top hit species distribution based on the number of BLAST top hits.

Peptides identified by Mascot were further validated by determining similarity with known proteins in the NCBI database using protein BLAST-P based on E-value generated and % similarity. All the sequences showed homology with protein/proteins from the databases. Thirty-four proteins were identified from 12 spots. The number of proteins identified was higher than the number of spots, due to 8 out of 12 spots (except spot number 2, 3, 4, and 12) align with more than one single protein. In addition, protein spot 9 and 10 were split from a combined spot but were excised from different gels due to their highly strong density. This experimental alteration generated almost a similar profile for proteins with a slight difference in these two spots in which leucoanthocyanidin reductase 1 was only found in protein 9 and elongation factor G family protein was only identified in protein 10.

The annotation of protein function and their cellular location is imperative to understand the role of proteins at the molecular level (Paz et al. 2017), and therefore, the identified proteins were subjected to Blast2GO analysis. Interestingly, some proteins fell into more than one GO category, depending on their (multiple) GO annotation. In terms of biological processes classification, the highest number of protein were associated with cellular processes (28.29%); metabolic process (23.24%); biological regulation (9.9%); regulation of biological process (9.9%); localization (8.8%); response to stimulus (8.8%); negative regulation of biological process (5.5%); and signalling (5.5%) (Figure 6.3).

By predicting subcellular localization, the highest number of protein were localized in cells (22%); cell parts (22%), followed by organelles (14%); membranes (12%); protein-containing complexes (11%); organelle parts (9%); membrane parts (8%), and extracellular regions (2%) (Figure 6.3). The identified proteins with molecular biochemical functions were classified into the following 5 categories according to their functions: the highest number of GO terms related to binding (44%) followed by catalytic activity (35%); transporter activity (7.15%); signal transducer activity (3.6%).

Among spots that were differentially expressed in BL, compared to NL, 12 spots with defined accession IDs were successfully retrieved after searching UniProt and NCBI database using the MultiIdent tool in ExPASy web server and BLAST, respectively. All the accessions were matched with other families rather than just the Celastraceae such as legume family (27.8%), edible vegetable plants of Brassicaceae family (11.1%), orchard species of Rosaceae family (11.1%), rice species (11.1%), and other species (38.9%). Interestingly, of those top hit species (by the number of BLAST top hits), 72.2% belonged to the Rosid clade (Figure 6.4). The results reflected the close phylogeny of family Fabaceae and Celastraceae in the taxonomy of Fabid (Rosid clade).

6.3.3 Analyses of differential expression protein

The protein expression pattern of *C. hindsii* in different phenotypes of NL and BL were analysed by 2DE. Of the 34 identified proteins from 12 spots, 12 were up-regulated (UR); 14 were only present (OP) in BL; and others were present in both (PB) (Appendix 4, Figure 6.2). UR and OP proteins involved in significant biological processes were further biochemically classified, and the classes of the proteins participating in mechanisms corresponding to plant growth and secondary metabolite and enzymatic variation between NL and BL could be revealed.

6.4 Discussion

6.4.1 Protein identification

A comprehensive strategy to identify proteins was developed in this study. Proteins from 6 spots (2, 3, 4, 9, 10, and 11) were identified by searching against MSPnr100 databases. The other 6 proteins were not recognised in this database but generated a range of peptides. The result was probably due to poor MS data, and perhaps the genome and

proteome of medicinal plant *C. hindsii* or even the family Celastraceae have been known very little in the literature. Further identification was validated by searching in the SWISS-PROT and the TrEMBL database via the web server ExPASy in three cases of spot 3, 4, 9, 10 and 11. These proteins are probably highly conserved across species, while some specific proteins are only present in relatively limited species (e.g. the regulatory protein of secondary metabolite synthesis). According to Wilkins & Williams (1997), the confident protein identification using mass fingerprinting data should be obtained from the first rank matchings whose scores are less than 30 and their ranks are far different the sequential matchings. For example, protein 2 was identified as acidic endochitinase, which found several matchings in MSPnr100 databases (Appendix 4). To make sure a confident identification of this protein, the MS data coupled with peptide mass fingerprinting data were used to search for homology against the SWISS-PROT and the TrEMBL database. All results (top 20 proteins generated in the list) confirmed the identification of chlorophyll a-b binding protein (all scores < 18) (Appendix 4). However, using this method coupled with MS/MS data to identify other proteins, the confident identification was not achieved indicated by the very high scores (high scores means less matching). This might be explained by Wilkins & Williams (1997) suggesting that cross-species protein identification by peptide mass fingerprinting will be useful only where phylogenetic distances between the species under study and those described in the database are small. The author also suggested that identity confidence will usually need to be increased through other techniques, which later confirmed by De Filippis & Magel (2012) who used mass fingerprinting data and BLAST search to identify proteins involved in the biosynthesis of flavonoids accumulated in the wood of *Robinia pseudoacacia.* On the other hand, the successful identification of the above 6 proteins may be due to their highly conserved evolution through the plant kingdom, which plays a key role in plant physiological processes as housekeeping proteins.

BLAST search for protein identification has been widely applied in recent proteomics studies as it rapidly determines similarity with known proteins in the available database (Buts et al. 2014; Armengaud et al. 2014; Upendar 2017; Kumar 2017; Na Li 2018). In this study, all identified proteins (with peptides highlighted as bold red and $p < 0.05$) found by nano-LC/MS/MS were similar to ones, which in turn were searched in the NCBI database for their other homologous sequences using bioinformatics tool - BLAST program. Other peptides sequences (highlighted as bold red and $p < 0.05$) had not been identified by the MASCOT, were also subjected to BLAST search to find similar proteins. The top hits of these BLAST results had an identity and e-value approaching 100% and 0.0, respectively, indicated confidential matchings and identification (Table 6.2). Base on searches, only the highest scored identifications by Mascot and BLAST will be considered as possible identification (excluding uncharacterised and hypothetical proteins). The intermediate and lower scored proteins are also worth to discuss because their e-value and percentage of identity are significant (close to 0 and 100%, respectively). Spot number 9 and 10 had significant similar identifications as they were split from combined spots which might contain more than one identity. Therefore, some repeated identification will be omitted.

Table 6.3 Proteins classified as high score, intermediate score and low score based on Mascot and BLAST search. Proteins with MS/MS scores from 60 and BLAST scores from 500 were categorised as high score; intermediate scores for those above one of either cut-offs; low score for those below both cut-offs. pI and MW are given as a comparison between gel expressed data and databases. Molecular weights were searched against the Swiss-Prot and TrEMBL (*Italics*) databases. pI values are theoretical and retrieved from MSPnr100 database by the MASCOT (protein 2, 3, 4, 9, 10, 11) and the rest of proteins from ExPASy via compute pI/Mw tool. Protein names in *Italics* are not likely matching as their high pI differences.

Pro	Protein name	Database		Gel	
		pI	MW (kDa)	pI	MW (KDa)
	High score				
2	Acidic endochitinase	4.5-6.6	13-36	4.7	20
3	ATP synthase CF1 alpha subunit	5.1-5.3	53-56	5.6	56
4	Chlorophyll a-b binding protein	5.1-6.0	12-36	5.2	20
10	Pentatricopeptide repeat-containing	6.0	20-162	5.1	20
	S locus-linked F box protein	6.0	*30-47*		
	Intermediate score				
9	Kunitz trypsin protease inhibitor	4.6	15-24	5.1	20
	Cullin-4	8.1	11-135		
	Leucoanthocyanidin reductase 1	5.4	45-52		
10	*Disease resistance protein RPS4-like*	8.3	28-76	5.1	20
	Germin-like protein 11	6.9	24		
	Elongation factor G family protein	8.9	12-87		
11	ATP synthase subunit d	5.4	19.5	5.2	20
12	Protein trichome birefringence-like	6.8-9.6	16-63	5.4	28.5
	Low score				
1	*Inorganic phosphate transporter*	8.0-9.6	56-61	4.8	16.2
	F-box/kelch-repeat protein	6.0-8.8	24-47		
	IQ-DOMAIN 1-like	10.4	50		
5	F-box/FBD/LRR-repeat protein	5.6-8.9	48-57	4.7	30
	Beta-1,3-galatosyltransferase 14	8.9	39		
6	Retrovirus-related Pol polyproteins from transposon 297 family	*6.3-9.4*	*7-158*	4.7	30
7	*Myosin-2-like isoform X1*	*8.1-8.9*	*124-137*	5.3	13
8	Endoribonuclease Dicer homolog	5.9-6.7	154-213	5.7	16.5
	Staphylococcal Nuclease domain-containing protein 1	*6.3-10.6*	*13-117*		
11	Auxin-binding protein ABP19a	6.25	22	5.2	20
	Autophagy-related protein 8D-like	8.6	13-14		

In short, comparative proteomic analysis using 2D-PAGE gels coupled with MS/MS technique revealed that the proteomes of NL and BL leaf phenotypes of *C. hindsii* were different. It was evident that the BL contains specific proteins not present in NL (Figure 6.2). Classification of *C.hindsii* proteins according to their molecular functions showed that nearly one-half of identified proteins were assigned to binding activity, particularly protein binding, RNA and nucleotide binding function, closely followed by catalytic activities. This agreed with proteomes of the medicinal plant *Opuntia spp* where the majority of protein function were also catalytic and protein binding (Pichereaux et al. 2015). The results reflected a strong activity of amino acid metabolism and nuclease metabolism in which proteins change their spatial and chemical structures to ensure probable and efficient function in plant tissues and respond to stimuli such as environmental stresses, fungal, bacteria and viruses. A range of enzymatic proteins was recorded with roles in secondary metabolite biosynthesis, post-translational modification, stress and defence response, and energy synthesis whose cellular locations mostly occurred in cell and cell parts (organelles). Discussion of these differentially expresses proteins will provide in detail to morphological and biochemical differences between NL and BL of *C. hindsii*. This differential display analysis could reveal the variation of growth and development and secondary metabolite accumulation between NL and BL.

6.4.2 Proteins facilitating growth and development through amino acid and nuclease metabolism

Plant proteins are an integral part of the chemical machinery for cell growth, division, and differentiation in young developing cells. As the most important compound of living organisms, proteins are well-known as their mechanical role which provides a structured framework, or scaffold, for nucleic acids and other cell components (Day 1996; Müntz 2007). More specifically, proteins play key roles in signalling transduction that either turn on or turn off the expression of genes in the nucleus by binding to regions of nucleic acid that have a regulatory function (Sjuts, Soll and Bölter, 2017; Su et al., 2018). Therefore, any change in content and structure of proteins in cells probably lead to changes in growth and development of plants, especially when they do not have an immunological system so as their survival is greatly affected by biotic and abiotic conditions. Results from both biochemical investigation (detailed in Chapter 5 – Enzymes and Metabolites) and the current proteomic study showed that there was a significant elevation of proteins BL of *C. hindsii*. These up-regulated proteins participate in diverse biological processes and might contribute to the superior growth and development of BL.

a. Proteins involved in amino acid turnover

In plant growth and development, proteins involved in various cellular activities such as cell division and differentiation, protein trafficking, stress, and hormonal responses (Naeem et al. 2015). The regulation of these processes required several key signalling mechanisms and the ultimate mechanism that ensures the quality of intracellular proteins is the selective destruction of misfolded or damaged polypeptides. This is because

misfolded proteins can arise in cells by mutation and through the various post-synthetic events, and as a result, they can be highly toxic as they tend to form intracellular aggregates whose accumulation would interfere with normal cell function and viability (Goldberg 2003). In addition, protein degradation supplies a major source of amino acids, which can be used for protein synthesis. The breakdown of proteins can take place in some compartments by molecular-specific mechanism though proteolytic enzymes or protein degradation complexes, such as the 26S proteome. Most of the proteins destined for degradation are labelled first by ubiquitin and then digested to small peptides by the large proteolytic complex, the 26S proteome with the participation of many gene products via sequential steps. This pathway is initiated with a conserved enzymatic cascade, consisting of E1, E2, and E3 enzymes, that are dedicated to attaching ubiquitin to selected protein (Naeem et al. 2015). This process is highly selective and precisely regulated by the E3 ligase, which binds to the specific substrate and catalyses the transfer the ubiquitin chain to the target protein (Guo et al. 2013).

* F-Box proteins and cullin-4 (upregulated in multiple spots)

Among E3 ligase, the Cullin-RING Ubiquitin ligases (CRL) represent the largest family in *Arabidopsis* (Hua & Vierstra 2011). In this complex, cullin protein acts as a scaffold, F-box protein as the substrate receptor, which is responsible for substrate specificity and defines the functional identity. F-box proteins contain a conserved F-box domain (40–50 amino acid) at their N-terminus (Bai et al. 1996). The C terminus of F-box proteins generally contains one or several highly variable protein-protein interaction domains, for example, Leu-rich repeat (LRR), Kelch repeat, tetratricopeptide repeat (TPR), and WD40 repeat that interact with specific targets. For example, Cullin 4 (CUL4) uses DDB1 (Damaged DNA binding protein 1) and DWD proteins (a subgroup of WD40 repeat proteins) as adaptors and substrate receptors, respectively, to assemble CRL4 E3 complexes (Hua & Vierstra 2011).

In *C. hindsii*, S locus-linked F-box protein, F-box/FBD/LRR-repeat protein At1g70590, and Cullin-4 were found up-regulated in BL. These three proteins are all important components of CRL system and are all up-regulated in BL. The result indicated that protein degradation by the ubiquitin-proteasome system extensively occurred and the plant would use that to generate adaptive responses in coping with various environmental stresses. Protein degradation is a common post-translational regulatory process in multi-cellular organisms to control divergent phases of almost all metabolic pathways. Recent studies from several groups have revealed that Cullin4 E3 ligase can target histones for ubiquitination, and importantly, histone ubiquitination may facilitate the cellular response to DNA damage (Dai & Wang 2006). Upon UV irradiation, CUL4-DDB-Ring of Cullin 1 (ROC1) is recruited to the damage foci through the binding of DDB2 to the damaged DNA (El-Mahdy et al. 2006). Similarly, Guo *et al.*, (2013) reported that CRL4 played a key role in DNA repair and DNA damage responses to genotoxic stress. In addition, a recent study from Roodbarkelari et al. (2010) confirmed that Cullin-4 as a major regulator of endoreplication in *Arabidopsis* trichomes. While endoreplication has

been found to play a developmental role in controlling cell size and regulating cell morphology and an increase DNA content is typically associated with larger cells (Harashima & Schnittger 2010). As a result, an increased level of cullin-4 probably explains its role in the formation of broader thicker leaf blade in some variants of *C. hindsii.*

Clinically, cullin-4 associated E3 ligases have been identified to involve numerous human diseases, especially about multiple cancer types (Chen et al. 2015). E3 ligases further recognise numerous substrates to participate in a variety of cellular processes, such as DNA damage and repair, cell death and cell cycle progression (Liu et al. 2009; Sarikas, Hartmann & Pan 2011; Kerzendorfer et al. 2011). These clinical implications were further discussed in Chapter 7 - Discussion and Conclusions.

Another notable member of F-box protein, which has extensively been researched in recent studies, is Kelch containing F-box (KFB) proteins. This protein belongs to F-box family, one of the largest superfamily of proteins in the plant kingdom (Naeem et al. 2015). Among many C-terminal motifs in F-box proteins, Kelch is one of the most frequent. Although the abundance of F-box domains and Kelch repeat are common in most living organisms, these two domains are only found together in plant proteins. KFBs are the most common F-box C-terminal domain in plants, their roles of some have been well established in circadian clock and flowering time regulation (Zhang et al. 2017), and recently have been revealed in various phenomena in plant physiology such as the roles in plant growth and development, secondary metabolism, and defense.

KFBs were expressed up-regulated in BL (protein 1), possibly indicating their roles in differentiating NL and BL in term of leaf morphological characteristic and secondary metabolites biosynthesis. Some of the recent studies have highlighted the role of KFBs in different physiological processes for the growth and development of plants. In *Arabidopsis*, overexpression of CFK1 (a plant-specific Kelch repeat containing F-box protein) produced phenotypes with elongated hypocotyls, DNA content and increase ploidy (Miyazaki et al. 2011; Franciosini et al. 2013). KFBs are also involved in seed germination, flower development in chickpea, and response to drought and salinity. Chen et al. (2013) reported that KFB from rice plays a role in the regulation of leaf senescence, grain number, and seed size. At transcriptional levels, a KFB encoding gene in rice was found to modulate the level of cytokinin, the plant hormones with vital roles in regulating cellular division, and to generate a better plant architecture, increase panicle size and improved grain yield per plant (Li et al. 2011). The disturbance of cytokinin translocation, distribution, or signal perception in plant severely degrades their vascular formation and lignification (Zhang et al. 2014). It is demonstrated that KFBs regulate the growth and development of several plant organs via their activity in selective protein degradation, and possibly to be a mediator in leaf blade expansion *of C. hindsii*.

b. Proteins interfering nucleic acid metabolism

Protein spot 8 was upregulated in BL and was likely not present in NL. MASCOT and Multiident tool failed to search protein matching for this spot. However, BLAST tool searched several matchings in the NBCI database with significant e-value and identity. Apart from high scored protein (F-box protein At1g70590), staphylococcal nuclease domain-containing protein 1 and endoribonuclease Dicer homolog were identified with closely lower scores but better e-value (Table 6.3). Therefore, they can be considered as possible identification.

* Staphylococcal nuclease domain-containing protein (upregulated in spot 8)

Plants adapt to environmental conditions via the modulation of gene expression. However, the rate of gene transcription is not prompt enough to respond to stress challenges. Plants have evolved an alternative mechanism to overcome this limitation by the translational suppression of non-housekeeping proteins and the sequestration of unnecessary mRNA transcripts into cytoplasmic structures such as stress granules and processing bodies (Vanderweyde et al. 2013; Muench, Zhang and Dahodwala, 2012). The compartment of a specific mRNA can be conducted by RNA binding proteins upon the perception of a stress stimulus. It was suggested that both stress granules and processing bodies play a role in translation suppression, while stress granules can sequester pro-apoptotic proteins, thereby protecting cells from death, processing bodies also have function in mRNA decay bodies (Vanderweyde et al. 2013; Xu and Chua, 2011). Staphylococcal nuclease domain-containing protein (or Tudor staphylococcal nuclease (TSN) was recently characterised as a novel component of stress granules in animal (Zhu et al. 2013) and *Arabidopsis* (Yan et al. 2014). In animal systems, TSN functions have been extensively described as regulation of transcription, stimulation of pre-mRNA splicing, regulation of RNA silencing (as a component of the RNA-induce silencing complex), and cleavage of hyper-edited double-stranded RNA (Li et al. 2018). However, in plants, the molecular function of TSN remain poorly characterised, mostly was described in *Arabidopsis*. For instance, TSN was reported as an essential part of reproduction and embryogenesis (Sundström et al. 2009). In addition, TSN also promotes stress tolerance through the stabilization of stress-regulated mRNAs encoding secreted proteins (dit Frey et al. 2010); to modulate mRNA levels in gibberellin biosynthesis (Yan et al. 2014); the TSN encoding gene CAN1 was showed to correlate with different forms of programmed cell death (Leśniewicz et al. 2012); and functions in mRNA catabolism during stress as a positive regulator of mRNA decapping (Gutierrez-Beltran et al. 2015). In few other species, TSN plays a role as mRNA transporter in rice (C. Wang et al. 2008) or has a cytoprotective role in Norway spruce as it was cleaved and inactivated by metacaspase family proteases during stress-induced cell death (Sundström et al. 2009).

It is commonly agreed that TSN is conserved in all studies organisms, comprising a tandem repeat of four staphylococcal Nuclease (SN) domains followed by a Tudor and C-terminal partial SN domain suggested to act as interaction platforms for nucleic acids and proteins, respectively (Li et al. 2018). These two domains can interact physically with

various proteins; however, they are differentiated by nucleolytic activity and RNA binding of SN. TSN upregulation in BL of *C. hidsii* may indicate that its role in promoting stress tolerance or protect the tissues from cell death. On the other hand, TSN as a component of RNA-induced silencing complexes (RISCs), has been research thoroughly as the significant mediator of metastasis (Yoo et al. 2011; Gutierrez-Beltran et al. 2016; Rajasekaran et al. 2016; Ma et al. 2015; Jariwala et al. 2017; Shao et al. 2017; Yu et al. 2017; Blanco et al. 2011). Its therapeutic implication will be further discussed in Chapter 7 - Discussion and Conclusions.

* Endoribonuclease Dicer homolog 3a (upregulated in spot 8)

Another alternative strategy that plant responses to protect cells from invasion by nucleic acids of pathogens such as virus and bacteria through RNA interference (RNAi) mechanism in which double-stranded RNA (ds RNA) molecules silence the post-transcriptional expression of homologous target genes (Ozcan et al. 2015). Small interfering RNA (siRNA) and micro-RNAs (miRNAs) are the most well-known members of a family of non-coding RNAs (ncRNA) that affect and regulate gen, RNA and protein function against stimuli (Carrington 2003). siRNAs are synthesised by the breakage of dsRNA molecules into shorter fragments with perfect complementarity by the cleavage of the cytosolic enzyme Dicer. Meanwhile, Dicer cleaves the pre-miRNAs (short hairpin structure of ds RNA) into shorter double-stranded miRNA with imperfect complementarity (Meng et al. 2017). These two small RNA (sRNA) are recognised by Argonaute 2 (AGO2) and RNA-induced silencing complex (RISC), where one of the strands is degraded, and the other strand guides the AGO2-RISC complex to bind and block translation of target mRNAs. Because of providing efficiency in silencing the target gene expression in a sequence-specific manner, these two sRNAs became indispensable tools to study the function of single genes (Ozcan et al., 2015; Amjad et al., 2017).

Dicer is one of the core proteins within the RNAi and is highly conserved from archaea to eukaryote and expand largely in plants and animals with multiple members of the family (Jia et al. 2017). Interestingly, many organisms encode multiple Dicer members, and in distinct organisms, multiple Dicers are involved in the RNAi pathway (Fukudome & Fukuhara 2017). In plants, Dicer has been evolved to at least four Dicer-like protein (DCL1-4). Of those, DCL1 mainly produces 21 nucleotides (nt) miRNAs, and the product generated by DCL2, DCL3, and DCL4 belong to various classes of siRNAs that are 22, 24 and 21 nt in length, respectively (Komiya, 2017; Tworak et al., 2016). An experiment conducted in rice (*Oryza sativa*) revealed that both knockdown and loss of function of rice DCL4, the homolog of *Arabidopsis* DCL4, lead to vegetative growth abnormalities and severe developmental defects in spikelet identity (Liu et al. 2007). Similarly, the silence of DCL1 resulted in the down-regulation of miRNAs in *Aphis gossypii* (Ma et al. 2017). At transcriptional levels, Tworak et al. (2016) reported the existence of six putative Dicer-like genes in the *Medicago truncatula* genome, and they

are expressed throughout the plant and significantly induced in root nodules with nitrogen-fixing bacteria and upon pathogen infection. Many of candidate genes were up-regulated in response to tomato yellow leaf curl virus infection and abiotic stresses in *Solanum lycopersicum* (Bai et al. 2012). Dicer is an important endonuclease in the biogenesis of miRNAs, and showed upregulation in BL of *C. hindsii*, suggested its roles in the growth and development of this phenotype. It can be believed that the up-regulation of Dicer in *C. hindsii* leaf extract probably associated with the biosynthetic activity of sRNAs. Other studies also revealed that miRNAs involving in different metabolic process, particularly play a role in the regulation of essential oil pathways in *Mentha spp*, biosynthesis of gingerol in *Zingiber officinale* (Singh et al. 2016) and also regulate terpenoid backbone, isoquinoline alkaloids, and flavonoids biosynthesis in *Curcuma longa* (Singh & Sharma 2017). Therefore, the elevated level of Dicer and probable sRNAs might associate with the superior growth and elevated level of flavonoids in BL; and might contribute to the medicinal properties of this species.

sRNAs have been well-known as the regulator of gene expression by translational repression and transcriptional silencing in several medicinal plants and the cross-kingdoms transfer of plant-derived sRNA have been reported and successfully demonstrated (Pirŕ et al., 2016; Ražná et al., 2016). Therefore, the presence of this precious compounds in *C. hindsii* marked this plant as a potential candidate for new disease treatment method on humans. However, there are several challenges with si/miRNA-bases therapeutics related to off-target effects, efficacy, immune response and toxicity, and especially the delivery of sRNA to target tissues (Sala-Cirtog, Marian and Anghel, 2015; Ražná et al., 2016). These matters were further discussed in the concluding chapter to elucidate the possibility of medicinal plant sRNA therapeutics in humans.

* Retrovirus-related Pol polyprotein from transposon 297 family (upregulated in spot 6)

Transposable elements are widespread components of plant genomes and can comprise from over 14% of the *Arabidopsis* genome to over 80% of the maize genomes. These genetic elements can move through the genome using DNA transposons or RNA retrotransposons. Insertions close or inside genes can directly disrupt gene function or generate new functionalities including gene silencing (Galindo-González et al. 2017). We found that retrovirus-related pol polyprotein from transposon 297 family (encode retrotransposon) were differentially expressed in BL of *C. hindsii*. In other genomic and proteomics studies on plants, transposable elements were significantly differentially expressed in response to a diverse array of activating stimuli (Grandbastien, 2015; Traylor-Knowles *et al.*, 2017). Matsunaga et al. (2015) reported that retrotransposons and transposons are activated during heat stress and regulated by small interfering RNAs (siRNA). Retrotransposons were upregulated in *Quercus suber* by wounding (Rocheta et

al. 2012); elevated by fungal elicitors, wounding, viral inoculation in *Nicotiana tabacum* (Anca et al. 2014); upregulated in response to heat stress, aphid infection and salicylic acid (Voronova et al. 2014)). Although it was not certain that these actions also happened in *C. hinssii*, it can be believed that the presence and upregulation of retrovirus-related pol polyprotein from transposon 297 family, coupled with results from the previous phytochemical studies suggested that they may be important in defence system of *C. hindsii* in response to biotic and abiotic stresses.

* Pentatricopeptide repeats (PPR)-containing protein

Pentatricopeptide repeats (PPR) are of the most popular repeat domain proteins which are reported to play several functional roles in growth, development, and stress response in plants (Sharma and Pandey, 2016; Mandal et al.2018; Xing et al. 2018). All PPR proteins are nuclear encoded, whereas a substantial portion of them have been thought to target to either the mitochondria or the chloroplasts (Schmitz-Linneweber & Small 2008). PPRs play a key role in RNA metabolism including RNA processing, splicing, stability, editing, and translation. Therefore, it is not surprising that their absence often causes lethality (Manna 2015). PPR proteins are also associated with photosynthetic defects, aberrant leaf development, changes in leaf pigmentation, tolerance to inhibitors of different biosynthetic pathways (Barkan & Small 2014). Zhang et al. (2017)characterised *Arabidopsis* mutant with defective chloroplasts, *term pigment-defective muntant3 (pdm3*), which exhibits a distinct albino phenotype in leaves, eventually leading to *pdm3* seedling lethality under autotrophic growth condition. Overexpression of the mitochondrial PPR40 gene in *Arabidopsis* resulted in enhanced germination and superior plant growth in saline conditions. Reduced amount of hydrogen peroxide, diminished lipid peroxidation, and lower ascorbate peroxidase and superoxide dismutase activities accompanied salt tolerance, suggesting that PPR40 can diminish the generation of reactive oxygen species (ROS) by protecting plants via reducing oxidative damage during stress (Zsigmond et al. 2008). PPRs were found to be present in three spots and strongly upregulated in BL of *C. hindsii,* indicated their vital role in the healthier growth of BL phenotype, and probably related to the accumulation of pigments as dark purple leaf colour was only found in BL. The purple leaf colour in medicinal plant often associates with the accumulation of anthocyanins in plants, which also show to have anti-bacterial activity against some pathogens (Skaar et al. 2014). Therefore, it can be concluded that upregulated level of PPRs in BL facilitating this phenotype in growth and pigment accumulation, probably anthocyanins. Rapid methods for screening and identification of mutations in PPR genes that affect the viability of crops, therefore, have significant implications in agriculture.

6.4.3 Proteins facilitating growth and development through stress and defence response

Plants resist pathogen attacks both with preformed defences such as secondary antimicrobial compounds (proteins, peptides) and by inducing defence responses. Inducible defences can be activated upon recognition of general elicitors such as bacterial flagellin and even host cell fragments released by pathogen damage and accomplished by temporal and spatial changes in gene expression that cause changes in the physiological and biochemical reactions to aid in resistance to the pathogen (Martin et al. 2003). The expression of these genes can result in the production and accumulation of the pathogenesis-related protein (PR), which can be induced by components of the plant's self-defence mechanism.

The current study confirmed the presence of the protein components of the defense system in the leaf tissues of *C. hindsii*. These proteins form the first line of defence against different stress conditions and help to prevent an attack of different pathogens. Comparative analysis of the stress and defence protein category from *C. hindsii* leaf extracts showed the presence of 8 proteins, which are listed in (Appendix 4) along with their corresponding molecular activity. These highly abundant proteins including germin-like protein, Kunitz family trypsin ad protease inhibitor 2-like, acidic endochitinase, disease resistance protein (CC-NBS-LRR class) family, and protein IQ-DOMAIN 1 isoform X. were found up-regulated in BL of *C. hindsii.*

a. Germin-like proteins (upregulated in spot 9 and 10)

The most abundant protein group among stress and defense-related proteins in *C. hindsii* leaves were germin-like proteins (GLPs) which are elevated in BL. They are one of the two distinct groups of a protein related to wheat germins (the other one is germins proteins) (Freitas et al. 2017). Both germins and GLPs have been described as the archetypal members of the cupin superfamily (Dunwell et al. 2008). GLPs are found in plant kingdom and exhibit several enzymatic activities such as superoxide dismutase (SOD) and polyphenol oxidase. In rare cases, GLPs can also exhibit oxalate oxidase (OXO) activity (the key enzymatic activity of germins proteins). The most evident function of GLP is the role in plant defense, which is conferred by the enzymatic antioxidative activity of OXO and SOD, both of which produce H_2O_2, a mediator of oxidative stress and cellular signalling.

It has now become evident that almost all the GLPs studied to date possess SOD activity (Guevara-Olvera et al. 2012). Recently, the SOD activity has been confirmed by Yasmin *et al.*, (2015) that a significantly higher activity of heat resistant superoxide dismutase (SOD) was observed in the transgenic plants (*T1 Nicotiana tabacum cv. Samsun*) with almost a uniform expression of OsRGLP1 in leaves, stem, and roots, as compared to the wild-type. Similarly, and elevated H2O2 levels and SOD enzyme activity in and GLP (GmGLP10) overexpressing transgenic tobacco plants, suggesting that GmGLP10 functions as a positive regulator of resistance to *S. sclerotiorum* (a non-

host-specific and necrotrophic fungus causing a destructive disease in crop production) (Zhang et al. 2018).

Although true germins have exclusively been associated with OXO activity, several recent studies reported GLPs to have both SOD and OXO activity. The most recent study revealed that GLP encoding gene (CsGLPs) had oxalate oxidase (OxO) or superoxide dismutase (SOD) activities were overexpressed during bud sprouting, under biotic and abiotic stresses which further confirmed their involvements in tea plant development and stress response (Fu et al. 2018). A GLPs recombinant protein (RmGLP2) was determined with OXO activity in cultured cells of tobacco expressing RmGLP2, which exhibited significantly reduced oxalate levels (Sakamoto et al. 2015). Furthermore, the latex from *Thevetia peruviana* was identified as rich in plant defence proteins, including a 120 kDa cysteine peptidase (with structural characteristics similar to germin-like proteins) having proteolytic activity (de Freitas et al. 2016). However, this protein exhibited no OXO and SOD activity or antifungal effects. Polyphenol oxidase (PPOs) catalysing the oxygen-dependent oxidation of phenols to quinones are universally distributed in plants and are assumed involving plant defence against pests and pathogens. A GLP with high PPO activity was identified and characterised in *Satsuma mandarine*. The PPOs are widely distributed in plants, which play a role in the browning reaction. This activity could elucidate the browning occurred to leaves of *C. hindsii* even in an intact state without any physical damage.

b. Acidic endochitinase (upregulated in spot 2)

C. hindsii contained acidic plant endochitinase [EC 3.2.1.14], the only protein identified in upregulated spot 2, which play an important role as an abundant disease resistance protein (Kasprzewska, 2003; Moravčíková et al. 2017). Acid endochitinase belongs to chitinases group, is an enzyme which catalyses the hydrolysis of internal β-1, 4-glycosidic bonds of the N-acetyl-D glucosamine polymer chitin, a component of fungal cell walls and exoskeleton of insects (Unhelkar et al. 2017). The chitin fragments released after the chitinase enzymes activity are further detected and recognised by pattern recognition receptor to trigger the additional innate immunity against the fungal infection (Chandra et al. 2017; Rodriguez-Moreno et al. 2018). The involvement of acid endochitinase in the mechanism of defence against pathogens has been reported (Carstens, Vivier and Pretorius, 2003; Rathore *et al.*, 2015). Chintinases are induced in *Brassica napus* after infection with *Sclerotinia sclerotiorum* (Zhao et al. 2007). The level of chitinases in rice cultivars correlated with resistance to sheath blight pathogen *Rhizotonia solani* (Shrestha et al. 2008).

Endochitinase triggers the antagonistic pathogen defence pathway in maize and suppress fungal pathogen growth on maize leaves (Ray et al. 2016). However, chintinases may also be involved in plant developmental process such a cell elongation (Zhong et al. 2002), cold tolerance (Yeh et al. 2000). Furthermore, they are identified as part of nodule protection machinery against pathogens (Ribeiro et al. 2011; Santos et al. 2010; Graça et al. 2016). A higher level of acidic endochitinases in BL was probably as the results of the

upregulated level of genes encoding the protein or their extensive role of the plant cell elongation, particularly in leaves. Thus, with their important and unique roles and functions, the induction of chitinase enzymes can be a useful tool for crop protection against fungal diseases, particularly medicinal plant crops.

c. Disease resistance protein RPS4-like

Another stress and defence-related proteins identified in *C. hindsii* was disease resistance protein RPS4 which also showed upregulation in BL. Plants have evolved a sophisticated innate immune system to detect pathogens, in which plant disease resistance proteins (R) recognise pathogen proteins (effectors) in a highly specific manner (Su et al. 2018). R proteins are defined by their stereotypical multidomain structure: an N-terminal Toll–interleukin receptor (TIR) or coiled-coil (CC) domain, a central nucleotide-binding (NB) domain, and a C-terminal leucine-rich repeat (LRR) (Ve et al. 2015). Therefore, R proteins can be further subdivided into two groups according to whether they carry a coiled-coil (CC) domain (RPS2, RPM1, RPS5, Rx, Mi) or a Toll/interleukin-1 like receptor (TIR) domain (RPS4, RPP1, RPP5, N) at their N termini (McHale et al. 2006). Some R proteins are involved in plant-pathogen interactions, such as the *Arabidopsis* TIR-NBS-LRR (RPS4) confers resistance to *Pseudomonas syringae pv. tomato* DC3000 strain carrying *avrRps4* (Gassmann et al. 1999). One member of CC-NDS-LRR was the powdery mildew resistance protein PM3B identified in a medicinal plant *Corydalis cava* (Papaveraceae) extracts and confers *AvrPm3b*-dependent resistance (Yahiaoui et al. 2004). The genes encoding the class resistance proteins RPS4 and RRS1 are closely positioned next to each other in a head-to-head configuration. They function together to recognize the bacterial effectors AvrRps4 from *P. syringae* and PopP2 from *Ralstonia solanacearum* and an uncharacterized effector from the fungus *Colletotrichum higginsianum* (Narusaka et al., 2016; Narusaka et al. 2016). The absence of RPS4/RRS1 and amino acid change in RPS4 cause loss of their functions and resistance to pathogens in *Arabidopsis* thaliana (Narusaka et al. 2017). TIR-only protein RBA1 and NPR1-1 recognise a pathogen effector to regulate cell death in *Arabidopsis* (Nishimura et al. 2017; Xiaoxiao Zhang et al. 2017). Other members of NBS-LRR R gene group have been shown to confer to virus, fungi, bacteria, alphids, and even nematodes) (Zhang et al. 2004; Su et al. 2018; Meyers, Morgante and Michelmore, 2002).

C. hindsii leaf extract possessed several different small low molecular defence-related proteins, which mainly belong to pathogenesis-related proteins such as Kunitz family trypsin and protease inhibitor 2-like, Auxin-binding protein ABP19a, serine/threonine-protein kinase PIX13, and protein IQ-DOMAIN 1 isoform X1. Of those, Kunitz family trypsin and protease inhibitor was found to inhibit the trypsin-like enzymes from the midgut of lepidopteran pests, to be fungicidal to *Candida buinensis* and bacteriostatic agent to *Escherichia coli* (Da Silva Bezerra et al. 2016). As termed as a post-translationally modified protein, auxin-binding protein ABP19a was reported conserved a significant peptide sequence similarity with germin-like proteins and identified as

proteins with antioxidant enzymatic activity such as β-glucosidase, glutathione S-transferase, manganese superoxide dismutase, which is important to plants responding to stresses (Ohmiya et al. 1998). Auxin control plant growth, differentiation, structural organisation, and auxin-binding proteins also reported being associated with auxin imbalance produced during some fungal infections and the upregulation of exogenous selenium in maintaining auxin balance after metal stress (Sun et al. 2016). Krishnan et al. (2015) suggested that the serine/threonine protein kinase has a possible role in modulating antifungal defence response in the plant, particularly against tobacco rattle virus-based virus-induced gene silencing in the non-model plant *Piper colubrinum*. The presence of isoleucine-glutamine (IQ) domain with calmodulin binding function in cyclic nucleotide-gated channels is supposed to be involved in diverse physiological functions, such as plant growth, adaptation to elevated Ca2+ concentration and response to abiotic and biotic stress (Fischer et al. 2013; Bürstenbinder et al. 2017). It seems that the presence of a number of low molecular pant pathogenesis-related proteins in *C.hindsii* leaf extract, which possess antimicrobial, antifungal and antiviral property, could support their strong biological activities, particularly the healthier fitness of BL phenotype.

6.4.4 Other growth and development-related proteins

a. Photosynthetic proteins (upregulated in spot 3 and 4)

The photosystem II (PSII) is a multiunit pigment-protein complex located in the thylakoid membrane of higher plants. Its function involves capturing and converting light into chemical energy, which is used to oxidise water and reduce plastoquinone in the light reaction of photosynthesis (Myouga et al. 2018). Chlorophyll a-b binding protein of LHCII is the core of this complex, also termed antenna complexes (Myouga et al. 2018). The major function of the antenna system is to capture light energy and transfer excitation energy to the reaction centre of the PSII. Meanwhile, ATP synthase catalyses the formation of ATP from ADP and inorganic phosphate in the presence of proton gradient across the inner membrane of chloroplast (Stock et al. 2000; Wang et al. 2015). These two proteins were also identified as chloroplast proteins, which respond to stimuli such as temperature and salts (Wang et al. 2014). In the current study, Chlorophyll a-b binding protein of LHCII and ATP synthase CF1 alpha subunit (chloroplast) showed a strong increase in BL of *C. hindsii* (spot 3). This increase may associate with the higher photochemical reactions and energy required for maintaining the superior growth and development of BL which, showed higher enzymatic activities and nuclease activities for secondary metabolites biosynthesis and other metabolic processes. Furthermore, cultivation condition of BL supported this result as BL plants were more exhibited to direct sunlight compared to NL plants, which were also planted in fewer nutrients soil.

It is believed that the elevated photosynthetic activities in chloroplast were correlated with proteins demand (Sjuts, Soll & Bölter 2017). Importing proteins into chloroplast is crucial during plastid development of plant life. Although the protein demand is higher in young and fast dividing tissues, compared to adult and non-dividing cell parts, this

importing activity into plastids is developmentally regulated, especially when exposing to fluctuating environmental conditions such as temperature and light (Li and Teng, 2013; Grimaud et al. 2013). Therefore, plants must regulate their protein content within the chloroplast to ensure optimal function of processes such as photosynthesis (Sjuts, Soll & Bölter 2017). Yfc1 (chloroplast) with a role in transporting proteins into plastids was found upregulated in BL. This result was further supporting that ATP synthase and chlorophyll a-b binding protein of LHCII type were also upregulated in BL as these plants were found to cultivated in more sunlight exposed sites.

b. Inorganic phosphates (upregulated in spot 1)

Phosphorous is one of the mineral nutrient and is of particular importance because it plays a key role in many crucial processes such as heredity (DNA, RNA) cellular compartmentalization (membrane lipids), energy metabolism (ATP) and phosphorylation-based signalling mechanisms (Puga et al. 2017; Jouhet, Maréchal and Block, 2007). Although phosphorous is relatively abundant in the environment, inorganic phosphate (Pi-the preferential phosphorous form directly absorbed by the plants) can easily be considered as one of the least available plant macronutrients (Rausch & Bucher 2002). Furthermore, there is always a sharp concentration gradient of Pi in plant tissues and the soils. Therefore, the role of Pi transporter has become crucial (Nussaume 2011).

Findings from Karlsson et al. (2015) and Liu et al. (2016) showed that a group of Pi transporters in vascuolar membrane regulate cytoplasmic Pi homeostasis and are required for ATP synthesis, fitness, and plant growth in *Arabidopsis*. To cope with Pi limitation, plants have evolved Pi starvation rescue systems that increase Pi acquisition and use efficiency and protecting them from stress caused by Pi starvation. The centre of these rescue systems was highlighted by the role of Pi transporter and transcriptional control including sugar, hormones, calcium and microRNAs (Puga et al. 2017; Ham et al. 2018). The study by Liu et al. (2014) also showed that miRNAs are an effective way for plants to monitor the turnover of Pi transporters in the membrane system by modulating the functioning of the membrane-associated ubiquitin machinery. These findings were in agreement with Fukudome & Fukuhara (2017) that Pi seemed to play a role in differential regulation of Dicer (enzyme catalyses the small RNA synthesis). Along with dsRNA-binding protein partners, Pi in a cell might play a vital role in sorting substrate dsRNA of various lengths toward appropriate dicing enzymes in *Arabidopsis* and other plants (Nagano et al. 2014). Therefore, upregulated levels of Pi transporter proteins (spot 1) and Dicer (spot 8) identified in the current study might illustrate their integral roles in Pi homeostasis, nuclease metabolism, fitness and plant growth in BL of *C. hindsii*. Improving our knowledge of Pi acquisition and use by the plant will certainly have a positive effect on reducing the dependency on fertiliser supply for crop production, followed by reducing environmental pollution and minimising the cost of crops (Nussaume 2011; Raven et al. 2018).

c. Myosins II isoforms (upregulated in spot 7)

Myosins consist of a large family of motor proteins that cover the energy released by ATP hydrolysis into mechanical force to move cargo along actin filaments (Kurth et al. 2017). Their cellular functions were described as driving cytoplasmic streaming, actin organisation, and cell expansion (Madison & Nebenführ 2013). Therefore, they were considered as the powerful players whose participation supports plant growth, environmental responses, and defence against pathogens (Citovsky & Liu 2017). The enzymatic properties of different myosin motors were mostly characterised from *Arabidopsis* but also from tobacco. For example, a *myo11b2* mutant analysis in *Arabidopsis* was established and showed a pivotal role in cell expansion and plant growth (Ryan & Nebenführ 2017).

This study also concluded that the growth rates of the mutant plants correlated roughly with the speed of the motors, demonstrating a clear relationship between intracellular motility and plant growth. Angiosperms typically encode around 15 myosin genes that can be grouped into two families of myosin VIII and myosin XI motor, whereas myosin II isoforms have not been reported in plants (Citovsky & Liu 2017; Nebenführ & Dixit 2018). Myosin II is a major force-producing, the actin-based motor in mammalian non-muscle cells, where it plays important roles in a broad range of fundamental biological processes, including cytokinesis, cell migration, and epithelial barrier function (Beach & Hammer 2015). There have been few studies to date assessing either protein expression profiles or behaviour of the multiple myosin II isoforms during cancer cell migration. The suggestion was that there was a significant positive correlation between expression levels of myosin light chain kinase (which activates myosin II) and likelihood of disease recurrence and metastasis (Eddinger & Meer 2007; Betapudi, Licate & Egelhoff 2006; Hindman & Virginia 2015).

6.4.5 Proteins involving secondary metabolite biosynthetic and post-translational modification

a. Secondary metabolite biosynthesis

C. hindsii is rich in a range of secondary metabolites and antioxidants such sesquiterpene, triterpenes, alkaloids, flavonoids, vitamin C and vitamin E. By concentrating on the secondary metabolites biosynthetic pathways, two proteins involved in the manufactory of flavonoids (particularly proanthocyanidins) and phenylpropanoids were identified: Leucoanthocyanidin reductase 1 (protein spot 9), Kelch-repeat F-box/protein (protein spot 1), respectively. These two proteins were found up-regulated and expressed in 2D gels with strong density. Leucoanthocyanidin reductase 1, although, had lower score than other matchings from spot 9, it can be a confidential identification of this combined spot and showed better BLAST search parameters. Similarly, Kelch-repeat F-box/protein was confidentially identified as it had the best e-value and identity.

* Leucoanthocyanidin reductase 1 (upregulated in spot 9)

Proanthocyanidins (PAs, also known as condensed tannins), the second most abundant plant polyphenolic compounds after lignin, are oligomeric and polymeric end products of the flavonoid biosynthetic pathway (Dixon et al. 2005). Their presence in the fruits, bark, leaves, and seeds provides many the growing plants the mechanisms of protection against herbivores, insect pests, and microbial pathogens. PA was ubiquitous in the species in Caprifoliaceae, Celastraceae, and Theaceae (Jiang et al. 2017). They also give flavour and astringency to beverages, medicinal herb and are increasingly recognized as having beneficial effects on human health (Ali & Ghasemzadeh 2011). For instance, PAs have been associated with reduced risks of cardiovascular disease, cancer, and Alzheimer's disease, can improve nutrition, prevent bloat in ruminant animals, and enhance soil nitrogen retention (Liu et al. 2016). In addition to these functional roles, PAs have recently been suggested to be beneficial to humans by providing cancer chemopreventative effects, and also through neuroprotective activities (Liu et al. 2013). Their well-known roles are believed to be conferred by several molecular functions: metal chelating activity that results in severe limitation of bacterial growth; associate with and irreversibly precipitate proteins, which is responsible for the astringent taste that repels herbivores. Moreover, PAs can be oxidized to quinones (a powerful antibiotic) which can initiate cross-linking of cell walls to increase the strength of this physical barrier to pathogens, therefore, the presence of proanthocyanidins is also a major quality factor for medicinal plant crops (Dixon et al. 2005). Due to their importance in plants and human health, it would be useful to have a better understanding of their biosynthetic pathway that could be useful for predicting and controlling the quality of medicinal plants.

PAs are oligomers and polymers of flavan-3-ols, primarily (–)-epicatechin and (+)-catechin. In PAs biosynthetic pathway, leucoanthocyanidin reductase (LAR) has been shown to convert leucocyanidin to (+)-catechin. It is commonly agreed in recent studies that LAR is a positive regulator of PAs biosynthesis. The LAR enzymatic reaction was recently found as a subsequent step in dihydro- flavonol 4-reductase reactions with enzyme preparations from some major crop plants (Pfeiffer et al. 2006). Liu et al. (2016) reported that loss of function of LAR in the model legume Medicago truncatula leads unexpectedly to loss of soluble epicatechin-derived PAs, increased levels of insoluble PAs, and accumulation of 4β-(S-cysteinyl)-epicatechin. LARs promoted the biosynthesis of catechin monomers and inhibited their polymerisation in transgenic tobacco (Pfeiffer et al. 2006). Other studies also confirmed the controlling role in PAs biosynthesis at the transcriptional level. In cacao leaves, flowers, and seeds, LAR gene was abundantly expressed and well correlated with PA accumulation levels, suggesting their active roles in PA synthesis in which the overexpressing TcLAR had decreased amounts of anthocyanidins and increased PAs. Similarly, in a recent study, the transcript abundance LAR observed at the late stage of fruit development was paralleled with the high amounts of insoluble PAs (Pfeiffer et al. 2006).

The biosynthesis and accumulation of polyphenols are regulated by environmental conditions such a temperature (Crifò et al. 2011), illumination (Koyama et al. 2012), water (Quiroga et al. 2012) and ultraviolet (UV) irradiation (Wen et al. 2015). The later

author examined the effect of postharvest UV irradiation on flavanol polyphenol accumulation in the grape berry, and the results indicated that the flavanol polyphenol reached its highest value when exposed to UV-C irradiation and induced the transcription of LAR gene and its gene products. BL was collected for the current proteomic experiments mostly from organised cultivations (plant nurseries, commercial gardens) which were believed to provide a better nutrients condition and UV exposure. In contrast, NL material was only collected from home gardens where they were cultivated basically for specimen conservation under shadow and nutrients deficiency. These limited environmental and fertilising provision probably restricted the growth of the plant, particularly leaf expansion, and the formation of specialised metabolites such as flavonoids.

* Kelch-repeat containing F-box proteins (upregulated in spot 1)

Kelch-repeat containing F-box proteins (KBFs) have also been found extensively involved in secondary plant metabolism and plant defence via controlling the activity of the gateway enzymatic catalysts (Zhang et al. 2013). Secondary metabolites play a crucial role in plant defence system as the production of secondary metabolites help plants to adapt to the environment and to cope with diverse types of stresses. These metabolites are also termed as natural products, have been used by humans to produce pharmaceuticals, nutraceuticals and food additives.

Many factors involved in this metabolism in plants and the identification of these factors is important for the engineering of useful compounds (Yu et al. 2012). KBFs involvement has been recently elucidated in secondary metabolites production. Some families of KFBs have been found to participate in the regulation of phenylpropanoid biosynthetic activity in *Arabidopsis*, via physical interaction with phenylalanine ammonia-lyase (PAL) and chalcone synthase (CHS) that catalyse the first and the committed step in the phenylpropanoid biosynthesis (Zhang & Liu 2015). KFB-PAL interaction ultimately mediates the proteolytic turnover of PAL by ubiquitination pathway, thus regulating the production of phenylpropanoids such as anthocyanins, flavonoids, phenolic esters and lignin (Zhang et al. 2017). These data strongly pointed out that KFB as a proteolytic mediator is an indispensable component in the regulation of PAL and CHS by balancing the homeostasis of cellular concentration of CHS and the production of flavonoids upon environmental stimuli. In another study, Feder et al. (2015)identified a KBF coding gene, when expressed, negatively regulates naringenin chalcone accumulation in muskmelon fruit by shifting the metabolic flux toward general phenylpropanoid products. In contrast, we found the level of KBFs and flavonoid were both upregulated in the current study of *C. hindsii*.

The increased level of KBFs probably balanced the homeostasis of cellular CHS or PAL concentrations, which effectively catalyse their reactions in phenylpropanoid biosynthesis. On the other hand, the higher level of flavonoids in BL was though negatively regulated by KBFs; it may be facilitated by other environmental factors such as temperature, illumination, and water to compensate the reduction by KBFs. Even

within the central enzyme-regulated reactions, the recognised control of these enzymes can occur through several mechanisms such as product inhibition, transcriptional and translational regulation, posttranslational inactivation and proteolysis, enzyme organisation/subcellular compartmentation, and metabolite feedback regulation. However, the roles of these factors are not recognised in *C. hindsii* and are upon the scope of this study. Nevertheless, understanding the regulatory role of KBFs in secondary metabolism might prove useful for plant breeders in tailoring new cultivars.

b. Post-translational modification

* β-1,3-galactosyltransferase (upregulated in spot 5)

Plant cells are surrounded by cell walls mainly composed of polysaccharides and proteins which are the key interacting components of the fundamental processes. Cell wall proteins and peptides are important players in cell walls contributing to their assembly and their remodelling during development and in response to stresses (Canut et al. 2016). To fulfil their contribution to wall architecture and function, cell wall proteins have undergone post-translation modification by interacting with other molecules. The attachment of sugar residues is the most complicated co-or posttranslational modification that a protein can perform. This protein-modified event reaches beyond the genome and is regulated by factors (e.g. epigenetics) that differ greatly among cell types (membranes, cytoplasm, or nucleus) and species (bacteria to eukaryotes) (Cosgrove 2014). The biogenesis of peptide-linked oligosaccharides results in the formation of the sugar-amino acid bond which is greatly regulated by the cellular enzymatic machinery. The glycopeptide bonds can be arranged in quite distinct groups, and in many cases, more than one type of sugar-amino acid bond can occur in the same protein, depending on the availability of catalytic enzymes (Spiro 2002; Breton et al. 2006; Kobata & Amano 2005).

β-1,3-galactosyltransferase involves in the biosynthesis of N- and O-glycans on arabinogalactan-proteins (AGPs) (Qu et al. 2008). AGPs belongs to the hydroxyproline-rich glycoprotein (HRGP) superfamily which constitutes the most abundant and diverse group of cell wall glycoproteins. APGs are proposed to play essential roles in a variety of plant growth and development process, including cell expansion, cell division, reproductive development, somatic embryogenesis, xylem differentiation, abiotic stress responses, and hormone signalling pathway in *Arabidopsis* (Seifert & Roberts 2007; Ellis et al. 2010). Therefore, up-regulated level of β-1, 3-galactosyltransferase may play an important role in growth and development in BL of *C. hindsii*, particularly in cell expansion and division of the leaf, via accelerating AGPs biosynthesis. Because AGPs consists of approximately 90% of polysaccharides which largely dictate the molecular surface of the glycoprotein molecule, it is likely that this carbohydrate moiety playing a critical role in AGP function and their glycosylation is greatly regulated by the enzyme machinery in which β-1,3-galactosyltransferase enzymes will work coordinately to regulate the density, length, and sequence of the galactan chain (Tan et al. 2012; Basu et al. 2015).

Clinical studies on the role of β-1,3-galactosyltransferase in glycosylation have been extendedly conducted on humans (Pinho & Reis 2015; Amado et al. 1998; Cheung et al. 2016; Felner et al. 1998) and were further discussed in the final chapter Discussion and Conclusions.

6.5 Conclusion

To our knowledge, this work represents the first comparative proteomic investigation *C. hindsii* between NL and BL phenotypes. Although *C. hindsii* genome has not been reported, proteomics has proved to be a valuable tool to identify proteins and their differential accumulation between phenotypes. The changes in protein abundance revealed significant variations between NL and BL proteomes and a certain level of protein degradation, particularly in NL. 34 proteins were identified from 12 spots and discussed to highlight the metabolic changes between NL and BL proteomes, with regards to proteins associated with flavonoids and phenylpropanoids biosynthesis, posttranslational modification, nuclease metabolism, stress, and defence-related proteins. It can be concluded that BL phenotype, which was characterised by healthier external morphological features, have better medicinal properties. Apart from the upregulated level of flavonoids, BL possesses higher levels of some disease-related proteins and bioactive compounds with promising therapeutic and pharmaceutical applications such as staphylococcal nuclease domain-containing protein, Kunitz trypsin protease inhibitor, cullin-4, beta-1,3-galactosyltransferase 14, and endoribonuclease Dicer homolog. This study improves our understanding of the medicinal properties of *C. hindsii* and provides a future direction for conservation and cultivation of medicinal plant crops, which can produce higher agronomic and pharmaceutical properties.

CHAPTER 7

DISCUSSION AND CONCLUSIONS

7.1 Conservation genetics

7.1.1 Methodological consideration

The conservation of rare and endangered plant species requires an accurate estimation of both the level and distribution of genetic variation, as this information is imperative for optimising sampling strategies for conserving and managing plant genetic resources (Allendorf, Hohenlohe & Luikart 2010; Ellegren & Ellegren 2016; Jiang et al. 2018). As genetic diversity is maintained, the fitness of plant populations is likely to be ensured, leading to a healthy plant system which may exclude the risk of extinction. For crop populations, genetic diversity is more variable due to human-made interference and cultivated conditions, which can lead to inbreeding. Therefore, maintenance of genetic stability is one of the key factors to ensure a plant system is producing a complete set of normal metabolites and natural products for growth and development (Ellegren & Ellegren 2016; De Filippis 2018). The selection of healthy plant populations (phenotypes or cultivars) with high genetic stability and maximum capability for normal production of natural products is greatly enhanced by genetic information.

The tropical and sub-tropical forest are estimated to contain up to 50% of all plant species of the high endemic level, but also have a high rate of biodiversity loss (De Filippis 2018). Therefore, much more attention for conservation has been paid to these areas in attempts to protect endemic, rare and endangered species from further loss and the risk of extinction (Ellegren & Ellegren 2016). For a developing country like Vietnam, an effective protection of biodiversity, particularly for plant resources, is likely to be the combination of conservation and utilisation, in which economic benefit should not be eliminated from the local people, whose livelihood are basically relied on some of these resources. Based on these criteria, *C. fugax* and *C. hindsii* were selected for genetic diversity analysis to support a conservation and management plan, as well as a suitable selection for utilisation. *C. fugax* is one of the rarest cycad species recorded in 2010 IUCN Red List and has a critically endangered (CR) status. The species is known only from an area of several square kilometres of Phu Tho province in the north of Vietnam (Osborne et al. 2007; Osborne et al. 2010). Wild-type *C. hindsii* has been seldomly found in forests, although their cultivated populations are spread over the provinces, mostly in the north part of the country. Interestingly, the broad leaf phenotype (BL) with better growth was found both in the wild and selected for cultivation by local people as the BL types are claimed to have better medicinal properties. Meanwhile, the narrow leaf phenotype (NL) has been deposited in several home gardens for specimen conservation and probably for domestic use. In this study, the amount and activity of several groups of secondary metabolites were higher in BL compared to NL, indicating potentially superior phenotype for medicinal plant cultivation (detailed in Chapter 5 – Enzymes and Metabolites). This indication was further supported by proteomic analysis as BL, which exhibited an

upregulation of critical enzymes in flavonoids biosynthesis and several bioactive compounds; which have been validated with pharmaceutical studies (detailed in Chapter 6 - Proteomics).

RAPD and RAMP markers have been widely used to study the genetic diversity and population structure of various endangered cycads like *Cycas guizhouensis* (Xiao et al. 2004), *Cycas balansae* complex, and *Cycas seemannii* (Keppel, Lee & Hodgskiss (2002) and medicinal plant species, such as *Celastrus paniculatus* (Senapati et al. 2013), *Commiphora wightii* (Harish et al. 2014) *Clinacanthus nutans* (Ismail et al. 2016), *Lonicera japonica* (Fu et al. 2013; Wei et al. 2016), *Moringa oleifera* (Avila-Treviño et al. 2017); amongst other plants. The efficiency of these PCR-base maker systems for analysing diversity relies upon the extent of polymorphism detected by a large number of loci in the genome, especially with non-requirement for cloning and sequencing information (Neale & Kremer 2011). The disadvantages of RAPD-PCR markers are that they were dominant and poor in reproducibility. However, these can be compensated by the addition of microsatellites primers (also called RAM-PCR) which are claimed to able to discriminate heterozygous from homozygous individuals (Bardakci 2001). This study found RAMP-PCR are more specific than RAPD-PCR because the longer SSR based primers enable higher stringency amplification, therefore reduce problems with reproducibility, a common criticism against the low-stringency RAPD analyses (Yang et al. 1996).

7.1.2 Research limitations

Due to strict regulation of Australian Biosecurity (i.e. AQIS), there were only insignificant amounts of dry leaves of both species allowed into Australia, and these amounts were not enough for highly replicated biochemical and proteomic experiments, which often require significant amounts of materials. This became more difficult in the case of *C. fugax* with the large size of seeds not allowed into Australia, which are believed to contain pharmaceutical properties. In addition, time and resources within a Ph.D project restricted the scale of investigations, therefore only *C. hindsii* was selected for subsequent investigations of phytochemicals and protein profiles. With only materials under desiccated state (air dried) available to us, the study was unable to conduct experiments associated with enzyme activities (except proteases which are thermally stable proteins/enzymes), and several stable metabolites were chosen for biochemical analyses. Proteins from leaves of *C. hindsii* were also found to be degraded during extraction, especially in narrow leaf samples. Although several modifications (the addition of protease inhibitors) were made in extraction procedures to minimise protein degradation, a certain level of degraded proteins were present in the final purified extract. Several unexpected and uncontrolled technical failures occurred during proteomic experiments which were believed to affect the sensitivity of the mass spectrometer. Therefore, peptide acquisition and identification were probably limited in several cases (i.e. fewer protein spots than normal). Although the study has experienced some unforeseen and uncontrolled difficulties, several experimental modifications and research strategies were applied successfully, and we believe them to be effective in achieving the

majority of the study goals and objectives; including enough replications for statistical analyses.

7.1.3 Genetic diversity of a rare and endangered plant – C. fugax

Intra-population and inter-population genetic diversity of *C. fugax* and for that matter also *C. hindsii* were analysed with the utilisation of molecular markers (PCR-based markers). The G-statistic for differentiation (Gst), Nei genetic diversity (H) and Jost genetic differentiation (D) were used. The results showed that both species contained moderate genetic diversity, but consistently low Gst values, indicating a considerable amount of genetic differentiation between populations and low genetic variation within populations. These results are consistent with those of other studies conducted on plants with close phylogenetic relationships, such as *Cycas balansae* (Xiao & Gong 2006) and the legume *Flemingia macrophylla* (Heider et al. 2007) belonging to Fabaceae, a close taxon of Celastraceae.

Knowledge of intra-population and inter-population genetic variation within species, coupled with information about their reproductive biology, is very important when establishing any conservation and management program (Frankham, Ballou & Briscoe 2010; Allendorf, Hohenlohe & Luikart 2010) aimed at preserving genetic variation within and among populations. Also, genetic diversity information is necessary to understand the evolutionary history of a species and in the assessment of future risk to diversity. This information is especially crucial for prioritisation of sites and management choices for future conservation programs (Ellegren & Ellegren 2016). In cycads, low genetic variation within a population and relatively high genetic differentiation between populations are characteristics of the biology and evolution of the genera (Walters & Decker-Walters 1991). Therefore, as an urgent response to the conservation of critically endangered species like *C. fugax*, effective methods of sexual reproduction through cross-pollination should be planned for implementation. Further statistical methods included PCA, MDS, and cluster analysis, and all were supportive for a rescue plan, as they apparently showed variable genetical distance between populations, but particularly of note, the natural forest populations contained the highest genetic diversity.

It is often the case that variability in smaller populations of plants can appear to contain nearly the same amount of variation as is contained in larger populations, and therefore, our data has mirrored these findings. However, care should be taken in concluding the study since it is not known how many of those seedlings sampled would have eventually reached maturity (Frankham, Ballou & Briscoe 2010; Allendorf 2017). The findings provide evidence for a slower than normal evolutionary genetic drift in small fragmented populations. A recommendation that Nm (estimation of gene flow) should be larger than 0.5-0.6 to avoid inbreeding depression is consistent with this data, according to the findings of Schou et al. (2017). Genetic diversity at the species level (and at the population level) was consistently different and suggested that much more genetic polymorphism existed among population than within populations.

The risk of extinction now faced by the remaining population of *C. fugax* is compounded by the species own biological characteristics, such as restricted habitat, poor adaptation, little seed dispersal mechanisms, and slow growing. Therefore, these factors facilitate a higher rate of inbreeding and potentially may lead to the loss of genetic variation, accumulation of deleterious alleles and inbreeding depression (Wayne & Bradley 2016). Consequently, the adaptation of the wild populations is likely to be low against climatic extremes, pollutants, pests, diseases and exploitation (Frankham et al. 2010). However, Schou et al. (2017) have claimed that there is no evidence that populations in the wild suffered from inbreeding depression, and that catastrophes, over-exploitation and demographic or environmental factor are more apparent causes of extinction than inbreeding. In the agreement, Waller (2015) suggested that extinctions may be incorrectly attributed to non-genetic factors alone when it is the interaction between genetic and non-genetic factors.

7.1.4 Conservation genetics of a medicinal plant – C. hindsii

C. hindsii has recently been widely cultivated as the public recognised its pharmaceutical properties claimed by traditional medicine practices and partly supported by current scientific evidence. It has been demonstrated that *C. hindsii* contains diverse secondary metabolites, and studies have identified several bioactive compounds including proteins and post-translational modification products that benefit human health. The majority of these products are synthesised and regulated by genetic mechanisms, apart from control by biotic and abiotic stresses (Nakabayashi & Saito 2015; Kamiloglu et al. 2016). Therefore, the occurrence of any phenomenon, such as genetic variation, accumulation of deleterious alleles, and inbreeding depression would potentially adversely affect the biosynthesis of these natural products. Since *C. hindsii* is being depleted from the wild due to its use in indigenous medicines, information on genetic diversity and structure are essential for their conservation program. However other biological characteristics like reproductive mode, seed dispersal mechanism, and geographic range have a significant influence on the levels and distribution of genetic diversity of plant species.

Similar to *C. fugax*, natural forest populations of *C. hindsii* contained a higher percentage of polymorphism and appeared to be in a habitat that was conducive to maintaining diversity in the wild. This is because the wild populations are unlikely to be affected by extremes of biotic and abiotic stress thanks to the intact habitats and ecosystem balance of reserved areas. PCA, MDS, cluster analysis, and Mantel test (correlation between geographic distance and genetic distance) were consistent amongst RAPD and RAMP data and clearly showed moderate genetic differentiation between populations, indicated that this species would also benefit from controlled breeding and hybridisation.

The high genetic diversity at the species levels is due to many factors and geographic distribution, and geographic distribution is probably one of the most influential factors that impacts on genetic diversity of C. *hindsii*. This species has been widely distributed in the north of Vietnam and southern China, occurring in the wild and cultivated habitats. With a wide range of geographic distribution (Muoi et al. 2009; Nam et al. 2013) and

high genetic diversity at the species level, but low diversity at the population level, genetic diversity will further deteriorate as commercial plantings of *C. hindsii* practice vegetative and clonal propagation. It can be predicted that in the longer term, the low diversity at the population level, frequent harvesting before maturity and vegetative propagation will further reduce the genetic diversity at population and species levels and the loss of rare alleles (Schoettle et al. 2012; Ismail et al. 2016). The authors also suggested that following genetic diversity analysis, it is essential to compare the phytochemical profiling among populations and phenotypes to understand the relationship between genetics and phytochemical production of this medicinal plant: which is important to developing crops for medicinal purposes.

It was concluded that further rescue efforts on one or a few of the populations would be considered inadequate for either species, and all four populations (more populations if present and identified) of both species should be preserved, especially *C. fugax*, as all four sites are within a very narrow geographic district. In case of *C. hindsii,* because of its valuable medicinal properties and high demand, this species needs more attention toward more serious conservation, in which an *ex situ* conservation program appears to be the most suitable and efficient measure for long-term protection.

Genetic diversity information potentially provides conservationists and plant breeders with options to develop, through selection and breeding, new and more productive crops that are resistant to biotic and abiotic stresses (Baruah et al. 2017). Also, selected phytochemicals and defence systems which maintain a healthy defence system of crops are important criteria for selecting better cultivars for propagation. Narrow leaf and broad leaf phenotypes of *C. hindsii* which are distinguished by morphologically on the whole show high polymorphisms. Therefore, further studies on other vital metabolites may give more information to molecular markers for ‘elite’ cultivar selection in conservation and breeding programs.

7.2 Anti-oxidative metabolites as key traits for medicinal plants

Medicinal plants are globally recognised as precious resources of new vitamins and metabolites, which play a primary role in plant growth and development, but more commercially used in human disease treatment. These plants have been historically used in traditional medicinal around the world, and when their pharmaceutical importance validated scientifically, they are usually used as sources of direct therapeutic agents and destination for new synthetic drugs (Gurnani et al. 2014). The global demand for plants with pharmaceutical properties is widespread and increasing, and while harvesting from the wild is claimed to cause the loss of genetic diversity and habitat destruction, another option for medicinal plants are cropping systems. This may constitute the primary production mode to ease conflict between the decrease of natural resources and the development of market demand for medicinal plants (Hamilton 2004; Li et al. 2015; Chen et al. 2016). Medicinal plant crops also offer opportunities to overcome problems that are inherent in wild extracts, such as misidentification, genetic and phenotypic variability, variable toxic components and contaminants (Raina, Chand & Sharma 2011).

Technically, one of the most challenging tasks in medicinal plant crop development is to select a proper phenotype/genome from original populations which show high production at the intra- and inter-species level. These variations can cause quality differences between herbal medicines, and consequently different clinical efficacy (Chen et al. 2016). This study examined the differences in concentration and activity of some antioxidative metabolites in two phenotypes NL and BL of *C. hindsii*, and the results showed that BL is potentially a better selection for development as a medicinal plant crop.

Using various spectrophotometric measurements, the accumulation of several important secondary metabolites, particularly α-tocopherol, flavonoid, and phospholipid were higher in broad leaf tissues of *C. hindsii*. The accumulation of these compounds is suggested to result from a range of factors, which depend on species, tissues, and developmental stages of the plant as well as the types and intensity of biotic and abiotic stresses. These metabolites play a vital role in plant physiological processes to protect plants from stimuli, especially insects and pathogens but have precious pharmaceutical properties, which are widely used in prevention and treatment of various human diseases. Therefore, plants possessing these compounds in abundance are considered as valuable targets for traditional medicinal practices and contemporary medicine.

α-tocopherol (vitamin E) is an essential antioxidant of the plant in response to stresses as it supplies protection to membranes, mainly by deactivating photosynthesis-derived reactive oxygen species (ROS) and preventing the propagation of lipid peroxidation as a lipid peroxyl radical scavenger (Gill & Tuteja 2010; Miret & Munn 2015). The accumulation of α-tocopherol in BL leaf tissue of *C. hindsii* is equivalent to that in α-tocopherol-rich plants of *Pistacia lentiscus var. chia* (app. 0.0053%) (Kivcak & Akay 2005). α-tocopherol was found to be the most stable vitamin under all drying conditions (86.4 % retention during oven drying) (Saini et al. 2014). Leaves of *C. hindsii* have been used in traditional medicine in the dry form, and it is believed that α-tocopherol is retained better and might bring good medicinal efficacy. Recent clinical studies have claimed that α-tocopherol appear to be the most promising therapeutic agent, and has been shown to reduce oxidative stress in non-alcoholic fatty liver disease (Hadi, Vettor & Rossato 2018; Takahashi et al. 2015). In addition, catechin-derived polyphenol, which was found in green tea and leaf extract of *C. hindsii* (catechin results in the Proteomics Chapter), was demonstrated to attenuate hepatic steatosis by decreasing adipose lipogenesis and enhancing hepatic antioxidant defences (Takahashi et al. 2015). These results might explain the reasons why *C. hindsii* dried tissues of BL have been used in traditional medicine to cure several liver-related diseases.

Flavonoids are one of the most studied secondary metabolites as they are the main target in many pharmaceutical studies of medicinal plants. Flavonoids are naturally occurring polyphenols that are ubiquitous in plant-based food such as fruits, vegetables, and teas, as well as in most medicinally used plants. There are over 10,000 flavonoids that have been characterised, and many have been proven to contribute to the chemoprevention of various types of cancers, such as leukaemia (Russo et al. 2017) and colorectal cancer (Cho

et al. 2017). Chemoprevention through diet modification, i.e., increased consumption of plant-based medicines, has emerged as a most promising and potentially cost-effective approach to reducing the risk of cancer (George et al. 2017). Flavonoid-rich diets have been studied as new sources of natural antioxidants which showed the potential to mitigate DNA damage induced by ROS and stimulate programmed deaths in cancer cells; thereby prevent mutations and consequently restore genomic stability (George, Dellaire & Rupasinghe 2017; Zamora-Ros et al. 2016; Ivey et al. 2017). These authors also suggested that dose selection of chemotherapeutic drugs is critical for therapy since some of the anticancer drugs possess systemic toxic effects when used at higher concentration. Taken all together, it may be concluded that BL of *C. hindsii* with an accumulation of flavonoids may be considered a good selection for development of a medicinal plant crop.

The presence and accumulation of proline, glutathione, and possibly phospholipids in BL of *C. hindsii* are also believed to be involved in antioxidants against ROS and oxidative damage, especially in biomembranes where phospholipids are important. However, the accumulation of these antioxidants not only respond to oxidative stresses but are also induce other factors, such as phytohormone levels or extreme stress. For examples, plant hormones which regulate major aspects of plant growth and development increased proline content in response to salinity (Tuna et al. 2008), whereas *Zea mays* and *Anabaena* plants treated by gibberellic acid (GA) reduced proline accumulation with salinity (Tuna et al. 2008; Alia & Gahiza 2007). The concentration of cellular GHS varies significantly under abiotic stress and may contribute to antioxidant function. While the concentration of GSH increases in plant exposed to drought (Samuilov et al. 2016) and heat stress (Wujeska-Klause et al. 2015), cold pre-treatment caused a slight decrease in the GHS pools of *Dunnaliella viridis* (Madadkar et al. 2014), and water deficit caused a reduction in GSH in seedlings of *Oryza sativa* (Pyngrope et al. 2013). In general, the presence of these antioxidants is crucial for normal plant life as they facilitate the adaptation of plant in response to changing conditions in the environment, particularly in crop plants as they help to maintain the stability and productivity of traits.

Protease activity, malondialdehyde (MDA) and protein content were also determined and compared between NL and BL of *C. hindsii.* Acid proteases (AP) were strongly elevated in NL compared to BL, while neutral protease (NP) activity was much lower overall and did not show differences between NL and BL. This pattern was also illustrated by 2D gels where most of the proteins migrated to the acidic side of the gels (detailed in Chapter 5 - Proteomics). These findings are in line with those from other medicinal plants where acid proteases were the major contributing factor for proteolytic activity (Fukayama, Abe & Uchida 2010; Bijina et al. 2011), while proteases within the neutral to alkaline regions were rarely found in plant (Rawlings et al. 2014).

Proteases are the main enzymes in protein degradation by different mechanisms after stressed-induce modification of proteins, and therefore their activity is claimed to increase in response to stresses, such as water stress (Simova-Stoilova et al. 2010), methyl jasmonate stress and mechanical wounding (Maciel et al. 2011). Therefore, the higher levels of AP in NL is possibly due to a strong response to stresses in which proteins were

facilitated to be degraded after leaves were detached from the plant and experienced desiccated; although this leaf preparation was the same for BL.

A decrease in protein content in NL possibly correlated with the higher activity of AP. This correlation was mirrored by the study of Kennedy & De Filippis (2004) where low protein content in leaf tissues was associated with the high proteolytic activity. It is important for plants to degrade proteins, which have been misfolded or damaged by mutation or various post-synthesis events, to so avoiding possible toxicity from intracellular aggregates (Goldberg 2003). However, for medicinal plants, the stability of proteins in plant tissues is important since many proteins and peptides have been found to be bioactive compounds and have therapeutic properties. Therefore, degradation of proteins might attenuate the quality of medicinal plants and thereby decrease their therapeutic efficacy. In addition to the higher activity of AP in NL, the amount of MDA was also found higher in NL, which further confirms that NL has undergone more stressful conditions, which might accelerate the generation of ROS causing lipid membrane peroxidation. This means that BL may have better protection mechanisms for leaf tissue membranes. It also highlights the role of phospholipids as antioxidants, which are also involved in membrane protection to reduce the production of MDA (Lin et al. 2014).

BL of *C. hindsii* appears to be the phenotype of value for development in medicinal cultivation since phytochemical assays showed higher content and activity of many secondary metabolites including flavonoids, tocopherol, phospholipid, and glutathione. Proteomic analysis also supported phytochemical assays as several enzymes were found to correlate with the upregulation of secondary metabolites, particularly flavonoids. Through this high throughput protein analysis, a number of proteins and peptides were identified and supported BL of *C. hindsii* as the better phenotype to select for drug cultivation.

7.3 Proteomics

7.3.1. Regulatory proteins in biosynthesis of secondary metabolites

The current proteomic study was conducted with the aim of identifying the cellular mechanisms of several valuable secondary metabolites and found that leucoanthocyanidin reductase (LAR) which regulates the production of proanthocyanidins (PA, a product of flavonoid biosynthesis) was upregulated in BL of *C. hindsii.* This finding matches those observed in earlier work of this study as flavonoid levels were significantly higher in BL (roughly 1.5 fold). Therefore, it is believed that the upregulation of LAR was a key factor responsible for the elevated accumulation of flavonoids in BL. These results also mirror those of previous studies that LARs promoted the biosynthesis of PAs in transgenic tobacco (Wang et al. 2018), cacao leaves (Shi et al. 2018), and the legume *Medicago truncatula* (Liu et al. 2016).

Another protein which has been demonstrated to be involved in phenylpropanoid biosynthesis and regulation were Kelch-repeat containing F-box proteins (KBFs). Although KBFs do not have an enzymatic function in biosynthetic pathways, their involvement was elucidated as a mediator in the turnover of phenylalanine ammonia-lyase (PAL) and chalcone synthase (CHS), two key enzymes catalysing the biosynthesis of phenylpropanoids. Recent studies have found that KBFs negatively regulated flavonoid and phenylpropanoid biosynthesis in *Arabidopsis* (Zhang et al. 2015; Zhang et al. 2017; Zhang & Liu 2015) and naringerin chalcone accumulation in muskmelon fruits (Feder et al. 2015). However, the findings of the current study did not support the previous research as KBFs were upregulated in higher flavonoid-accumulated leaf tissues of BL. A possible explanation for these results may be due to the complexity of secondary metabolite synthesis and regulation of enzymes including PAL and GH, which are finely regulated by multiple genes in a feedback system; some of which are only expressed in specific tissues or under certain environmental conditions (Tiago et al. 2017). Another possible explanation for this is that our studies were conducted in *in vitro* conditions, whereas secondary metabolites accumulation in *C. hindsii* is regulated by various environmental conditions *in vivo*, which may lead to contrary results. In *in vivo* conditions, secondary plant metabolites, particularly flavonoids, are greatly influenced by temperature, water availability, and UV irradiation which are believed to be the key facilitating mediators of flavonoids biosynthesis (Zandalinas et al. 2017). KBFs should be considered as a potential candidate in anti-sense technological studies with the aim to increase the concentration of secondary metabolites in plant tissues of *C. hindsii*. The current findings not only confirmed that BL phenotype is a better source of breeding material for medicinal crop development but may lead to plant breeders in tailoring a selected cultivar by growing them in controlled environmental conditions.

7.3.2. Production of valuable secondary metabolites

The quality of medicinal plants is substantially determined by the composition and concentration of relevant natural products. These products may respond to treatment with various elicitors, signal compounds and abiotic stresses. In the field, it is well-known that plant response to synthesise and accumulate specific secondary metabolites is likely to be affected by a combination of different stresses rather than by a single factor. To improve crop quality of medicinal plants, it is essential to understand which kind of stress condition enhances the accumulation of the target compounds without inducing losses in biomass or general plant health.

The combination of drought and heat stress is considered one of the most chronic conditions that take place in natural environments, affecting plant growth and productivity, leading to a specific regulation of the metabolome depending on the particular stress and the plant species. There have been many reports to show the effect of water deficit and temperature on the accumulation of diverse secondary metabolites in commercially used species, indicating a putative high economical relevance of these findings (Kleinwächter & Selmar (2014); Kleinwächter et al. (2015); Sancho-Knapik et al. (2017); Tiago et al. (2017) and Zandalinas et al. (2017)). Therefore, in case of *C. hindsii*, it is important to undergo experiments to test whether drought/temperature-induced

stresses enhance or inhibit the accumulation of certain metabolites, particularly flavonoids, and to what level of water deficit the accumulation is optimum, because there has been limited information on how the metabolome of C. *hindsii* or species in family Celastraceae respond to stresses. Under field conditions for medicinal plant crops, it is hard to control the environmental conditions due to the high cost of the controlling system. However, it can be done through the selection of sites for cultivation where climate condition is close to the optimal tested conditions. In Vietnam, the likelihood to select the suitable location is high as the country is spread in length from north to south direction including highland areas. Therefore, a wide diverse range of climates make more options for site selection available.

In cell culture environments, plant growth regulators (PGRs) and fungal elicitors induce the production of secondary metabolites in many medicinal plants (for review see Sancho-Knapik et al. 2017; Jamwal, Bhattacharya & Puri 2018), ranging from 1-2 fold in *Catharanthus roseus* to 28 fold in *Scutellaria baicalensis* (Yoon et al. 2000). Flavonoid content has been shown to increase with PGR treatment with different combinations and concentrations of cytokinins and auxins, strongly stimulating the production of flavonoids in cell suspension cultures of *Digitalis lanata* (Bota & Deliu 2015). Total flavonoids production in *Orthosiphon stamineus* were influenced by levels of abscisic acid (ABA) (Ibrahim & Jaafar 2013). It was also observed that ABA enhanced flavonoid production in *Thymus vulgaris* (Ibrahim & Jaafar 2013). Stress signal molecule-based elicitors of secondary metabolite biosynthesis have frequently been used successfully in plant cell and tissue culture (for review see Ramakrishna & Ravishankar (2011). Kleinwächter et al. (2015) reported that the impact of MeJa applications varied depending on plant species and this stress signal transducer strongly enhanced the concentrations of benzyl isoquinoline alkaloids in *Chelidonium majus* (+46%) and flavones in *Petroselium crispum* (+70%). PGRs were used in tissue culture for clonal propagation of *C. hindsii* in Vietnam as stimulants and growth substances; several medium formulations were demonstrated to increase the survival rate from 75% to 95% (Loan 2012; Nam, Thang & Tho 2013). Upon the successes of these studies are a good foundation to develop future projects targeting on the relationship of secondary metabolites and PGRs in tissue culture.

In contrast to using PGRs and elicitors, scientists have also used some inhibitors to block competing reactions such as the squalene synthase and β-caryophyllene synthase; the enzymes consume farnesyl diphosphate for sterol and β-caryophyllene biosynthesis (leading to the production of artemisinin – a bioactive compound to treat malaria) in *Artemisia annua*. The addition of these inhibitors has been proven to elevate artemisinin production by 3.14 and 5.49 fold, respectively. These finding cannot be entirely extrapolated to all other plants without further validation, particularly for *C. hindsii*, which is phylogenetic distance to *Artemisia*.

In short, biochemical integration has strongly enhanced the possible discovery and production of pharmaceutical and culinary products (Rai et al. 2017). Genomics, metabolomics and proteomics are high-throughput technologies that may help speed up the determination of the mode of action of phytomedicines and allows investigation of

herbal extracts prominent active principles. Although metabolomics and proteomics techniques have generally proven valuable, they still face substantial challenges, including clear large-scale metabolite identification. However, further development of the metabolomics field in general could provide better tools for the discovery of the next generation of natural products inspired by popular local knowledge, gathered in ethnobotanical studies and enhanced by recent molecular phylogenetic approaches (De Filippis 2018).

7.4 Potential for new bioactive compounds

By comparative proteomic analysis, the current study found some proteins regulating biosynthesis of some classes of secondary metabolites. This is also in accordance with previous observations, which showed that leaf extracts of *C. hindsii* contained a variety of important specialised secondary metabolites including vitamin C, vitamin E, sesquiterpene, triterpenes, alkaloids, and flavonoids (Hu et al. 2013, 2014; Ly, Shimoyamada & Yamauchi 2006; Yao-Haur Kuo, Chen & Kuo 1995). This study also found a number of proteins and peptides, which have been implicated in playing a critical role in 'state–of–the–art' advanced therapeutic intervention in human diseases. However, not all protein spots were confidentially identified by Mascot scores, and further searches for homologies were performed by UniProt (MultiIdent tool) and the NBCI (BLAST tool). This uncertain was probably because of the unavailability of protein information in the current databases and poorly generated MS/MS data. Therefore, two categories (high scores and low scores) of identified proteins based on their matchings against selected databases were identified and detailed in Table 7.1. For both groups, we found proteins correlated with medicinal properties. Therefore, it is worth briefly discussing their importance and implications associated with the modern development of pharmaceuticals and human disease therapies. It is interesting to note that most of all these gene products and proteins have been found upregulated in BL of *C. hindsii*, indicating that this phenotype is most probably a superior cultivar for medicinal plant crop system development.

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

Protein	Biochemical role	Human health effects	References
Kunitz - trypsin type proteinase inhibitor	Defensive role against biotic stress	- Inhibits trypsin-like enzymes and has some fungicidal activity. - Prevents the proliferation of colorectal adenocarcinoma, and B16F1 melanoma cells.	Da Silva et al. (2016) Price et al. (2016) Roy et al. (2018)
Dicer	Catalyses the biosynthesis of mi/siRNAs	- Associates with prominent and important types of cancers like breast cancer, leukaemia and lung cancer. - Deregulates miRNAs, and acts on invasion and metastasis in epithelial to mesenchymal transition (EMT) of cancer cells. - Have inhibitory effect on pancreatic cancer.	Kumar et al. 2017 Debnath et al. (2017) Mohammadi et al. (2017) Srivastava et al. 2011
β-1,3-galactosyl-transferase	Catalyses the biosynthesis of arabinogalactan-proteins	- Exhibits anti-proliferative effect on HeLa cells; blocks metastasis of liver tumour cells; induces apoptosis in breast cancer cells. - Presents an improvement in gastro-protective action. - Modulates the skin innate immune system. - Associated with immunoglobulin A nephropathy (IgAN) due to abnormalities of IgA1 O-glycosylation.	Bento et al. (2014) Moghtaderi et al. (2017) Scoparo et al. (2016) Zahid et al. (2017) Yeo et al. (2017)
Cullin-4	Controls cell cycle, DNA damage, DNA repair and apoptosis	Associated with reducing tumour size, and affects cell proliferation, migration, invasion, and cancer aggressiveness.	Cui et al. (2016) Jang et al. (2018) Li & Wang (2017)
Staphylococcal nuclease domain-containing protein	Transcriptional regulator via mRNA splicing and stability	- Contributes to cell proliferation, colony and tumour formation. -High expression in invasive/metastatic breast cancer cell lines compared to poor cancer lines. - Regulates malignant glioma invasion.	Yan et al. (2014) Blanco et al. (2011) Emdad et al. (2013 Jie et al. (2013)

7.4.1 Proteins with antitumor / anti-inflammatory activity and therapeutic applications

a. Kunitz-trypsin type proteinase inhibitor (high score proteins)

Kunitz-trypsin type proteinase inhibitor (KTI) has played a defensive role against biotic stress by limiting cell damage in diverse plant tissues, including seeds, tubers, leaves, rhizomes, and fruits (Tsukuda et al. 2006; Macedo et al. 2016). Plant-derived protease inhibitors have not widely been used commercially, but their pharmaceutical applications appear to have increased as control of plant disease by green plant bioproducts approach is currently one of the most dynamic areas of research in agricultural biotechnology. In addition, recent studies have revealed more pharmaceutical properties of plant-derived protease inhibitors, including KTI, which were identified in leaf extract of *C. hindsii* (Da Silva Bezerra et al. 2016; Price et al. 2016; Roy et al. 2018). Kunitz protease inhibitors including Kunitz-type serine protease inhibitor also show potential against invasive tumour growth and metastasis of human prostate cancer (Wu et al. 2017), and thyroid cancer (Liu et al. 2018). Since KTIs are stable over a range of temperatures and abundant in *C. hindsii*, it can be suggested that KTIs may play a role in anti-tumour and anti-inflammatory properties.

b. Dicer and its biogenesis product – mi/siRNAs (both low score protein)

Considered as a contemporary therapeutic target in clinical studies, the abundance of plant miRNA has been demonstrated to be successful in mammalian disease therapies, especially in cancer treatment. Several initial demonstrations of the cross-kingdom miRNA-regulated phenomenon was recorded by the studies by Zhang et al. (2012), Mlotshwa et al. (2015), and Chin et al. (2016), indicating that miRNAs can be acquired by mammalians through food intake, and can function on target genes to suppress the growth of tumours. As a result, the hypothesis was drawn that engineered edible plants producing mammalian tumour suppressor miRNAs might be an effective, nontoxic, and inexpensive chemo-preventive treatment strategy for human cancers. In humans, it was estimation that 60% of all protein-coding genes are targeted by miRNAs, affecting virtually every physiological process in the body. Therefore, a diverse array of human diseases is associated with dysregulation of miRNAs, particularly tumour suppressor miRNAs (Chakrabarti et al. 2013; Debnath et al. 2017; Del Cornò et al. 2017; Kumar et al. 2017; Mohammadi et al. 2017; Sala-Cirtog, Marian & Anghel 2015; Srivastava et al. 2011; Zhao et al. 2018).

Recent studies have addressed this controversy and confirmed that ingested miRNA and/or small non-coding RNAs from plant sources taken up by the digestive system, were delivered to the recipient cells; including Tudor staphylococcal nuclease proteins (Mlotshwa et al. 2015; Ražná et al. 2016; Otsuka et al. 2018; Zhao et al. 2017; Del Cornò et al. 2017). Such studies have generated considerable excitement because it raises the possibility of bioengineering medicinal plants to produce therapeutic miRNA that could

then be delivered to affected tissues by ingestion (Witwer & Hirschi 2014). The widespread cultivation and consumption of *C. hindsii* have the potential for miRNA transfer mechanisms to be evaluated and account for beneficial properties of products from this species.

c. β-1,3-galactosyltransferase and its biosynthetic product, arabinogalactan-proteins (both low score proteins)

Plant cell wall proteoglycans (AGPs) and are composed primarily of galactose and arabinose linked to proteins rich in hydroxyproline, serine, and threonine (Ndeh & Gilbert 2018). AGPs isolated from *Larix occidentalis*, *Panax ginseng*, *Coffea arabica* were capable of activating macrophages *in vitro* by augmenting nitric oxide, superoxide anion, and cytokine production (Moghtaderi, Sepehri & Attari 2017; Nosáľová et al. 2011). The purified AGPs from the *Endopleura uchi* bark exhibited anti-proliferative effect against HeLa cells (Bento et al. 2014) and blocked metastasis of liver tumour cells. An AGP from alkaline extracts of green and black teas was described and presented a significant gastro-protective action (Scoparo et al. 2016). Interestingly, AGPs from baobab and acacia seeds could modulate skin innate immune responses conferred by immunological and dermatological activities (Shakhtshneider et al. 2016; Zahid et al. 2017). A number of crude plant glycoproteins have been proven to be anti-inflammatory and anti-cancerous in mammalians via oral administration (Pinho & Reis 2015; Rozov, Permyakova & Deineko 2018), the possibility to develop medicinal crops for bioactive compounds extraction from raw materials (leaves, stems, roots, etc.) has become promising in attempting to produce precious health benefits, and it may also be true in the case of the medicinal plant *C. hindsii*.

It has become evident in recent years that defects in the attachment of carbohydrate to proteins have been implicated in a number of human diseases. The congenital disorders of glycosylation represent a group of systemic diseases characterised most prominently by neurological and developmental deficiencies, and these have been well described at a molecular level (Nakanishi & Yoshikawa 2016). Taken all together, β-1,3-galactosyltransferase and its biosynthetic products, arabinogalactan-proteins should be further studied to extend our understanding of their pharmaceutical properties.

7.4.2. Proteins overexpressed in tumour cells and their therapeutic applications

a. Cullin-4 associated E3 ligase (high score protein)

The ubiquitin-proteasome system promotes the timely degradation of short-lived proteins with key regulatory roles in a vast array of biological processes, such as cell cycle progression, oncogenesis, and genome integrity (Cui et al. 2016). Cullin-RING ubiquitin ligase 4 (CRL4) is implicated in controlling the cell cycle, DNA damage repair and apoptosis (Jang et al. 2018). CRL4 protein, including Cullin4A and Cullin4B are often accumulated in human malignancies. Overexpression of Cullin4A/B have been found in many cancers, and generally associated with tumour size, cell proliferation, migration,

invasion, and cancer aggressiveness (Jia et al. 2017; Li & Wang 2017; Jang, Redon & Aladjem 2018). The elevated CRL4 attenuated DNA damage repair, and increase genome instability is believed to facilitate tumorigenesis. Taken together, Cullin4 can influence human disease development and progression by regulating a wide range of cellular processes; and it can provide new insights into cancer diagnosis and treatment, especially the identification of CRL substrates, which is little known so far.

b. Staphylococcal nuclease domain-containing protein (low score protein) – biomarker in cancer prognostic

In plant systems, the molecular function of NAME IN FULL TSN has not been described thoroughly, most have been characterised in the model plant *Arabidopsis* as a transcriptional regulator and a mRNA splicing and stability factor in response to stresses (Yan et al. 2014). This is probably the first time in the literature, TSN has been reported to be accumulated in a medicinal plant like *C. hindsii,* and it might play a role in protecting the plant from stress-induced cell death. In contrast to plant systems, TSN has been researched thoroughly in mammalian systems, particularly in human beings as it has important roles in cancer (Li et al. 2018). TSN is a nuclease in NAME RISCs and is required for optimum RISC activities in facilitating small interfering RNA (siRNA) and micro RNA (miRNA) mediated silencing of some reporter gene (Jariwala et al. 2017; Shao et al. 2017).

TSN has housekeeping roles by stimulating transcription, mRNA splicing and silencing. It has become evident that its increased expression is closely associated with various types of cancers, indicating TSN is an attractive target for anti-cancer therapy and a potent tumour marker. Other evidence also suggested that TSN can be used as novel diagnostic and prognostic of prostate cancer (Kuruma et al. 2009), colorectal cancer (Zhu & Tan 2017), colon cancer (Wang et al. 2012), and lung cancer (Xing et al. 2018). The functional activation of TSN might be a novel mechanism underlying the early carcinogenesis of cancer which may be used to predict survival in cancer patients. In conclusion, the challenges in engineering TSN inhibitors in several studies suggested that plant-derived TSN proteins may be used as a diagnosis and prognostic biomarker in cancer therapy.

7.5 Plant propagation, breeding, and cultivation

7.5.1. Propagation and re-introduction of C. fugax

Plant propagation requires an understanding of the reproductive systems of the plants and is critical for the preservation and management of plant resources (Wilcock & Neiland 2002). This is especially true for critically endangered cycad species like *C. fugax,* whose breeding system is through cross-pollination between genetically distinct plants (xenogamy). Propagation of cycads from seed is essential for their conservation as this method appears to be the most feasible method which can help supply traditional and horticultural demand for plants and may reduce exploitation pressure on wild populations (Calonje et al. 2011). In addition, seed-propagated cycad individuals can also be

reintroduced to habitats to prevent the genetic erosion of wild populations (Walters 1999). However, parent cycads need to be well documented to avoid genetic contamination (e.g. unwanted hybridisation) from other cycads during the propagation process.

Since existing populations and individuals of *C. fugax* are separated in fragmented habitats, their reproduction often fails to achieve full pollination potential. The study survey observed that *C. fugax* are planted in small clusters of two individuals (or a few more) including both male and female (male and female plants can easily be distinguished morphologically by vegetative and reproductive organs, where male plants and cones are usually smaller than female one). Several observed populations include parents and their offsprings, indicating pollination between close relatives which might promote inbreeding depression and expression of harmful recessive alleles (Richardson et al. 2016). Other reproductive characteristics, such as the large size of seeds and possibly a lack of pollinators for outbreeding, may result in genetic drift and inbreeding of *C. fugax*. To rescue populations from these threats, conservation biologists can augment gene flow into small populations to increase variation and reduce inbreeding depression. There have been few cases in which inbred plant populations were rescued, as demonstrated in the re-introduction efforts of the Mexican cycad *Zamia furfuracea* (Donalson et al. 2003) and other plant species, such as chestnut (Wheeler & Sederoff 2008; Miller et al. 2014) and maybe willow (Mosner et al. 2012); although the latter has been poorly documented. Techniques required for successful seed propagation of cycads have been described, including pollen collection, storage, viability testing, manual pollination, seed collection, storage, and germination. These practices need to be tested if applicable to *C. fugax* in order to increase the availability of this cycad and reduce the demand for wild-collected plants.

7.5.2 Plant breeding for superior plants in C. hindsii

For commercial crop plants, there has been a continuous attempt to alter their genetic architecture for efficient utilisation of food, fibre, fuel, pharmaceutical substances or other end uses. Of those efforts, domestication and selection by humans of desirable traits have contributed a great deal to ensure increased production and may provide a solution to many of the problems associated with intensive agriculture (Ochatt & Jain 2007). A total of approximately 200 major crops and minor crop plants have been domesticated, and the achievements have led to changes in phenotype and progression from the wild to the cultivated forms, and different conservation status (Jain & Gupta 2013). Interestingly, many of recent domesticates have been produced by intensive agricultural research but not by gradual directed breeding methods.

Medicinal plants are classified as industrial crops, which provide special materials (essential oil, pharmaceuticals, colorants, dyes, cosmetics, and biocides) (Lubbe & Verpoorte 2011). However, most of them have not been domesticated and currently there is no ‘tool-kit’ to improve their medicinal attributes for better clinical efficacy. That is because successful domestication and improvement are unlikely to be achieved without deeper insights into the evolution pattern of the plant genome. Therefore, studies of

medicinal plant genome evolution are crucial not only for the understanding of the ubiquitous mechanisms of plant evolution and phylogeny but also for the sustainable utilisation of plant pharmaceutical resources (Kroc et al. 2014). *C. hindsii* is a semi-domesticated (actually poor for a commercial crop) medicinal plant, which has been chosen for cultivation based on traditional medicine knowledge but rarely by biomarkers (e.g. secondary metabolites and proteins) that are partially revealed in the current study. However, knowledge about the genome and the announcement of plant-based drug discovery and development of *C. hindsii* might bring this medicinal plant close to being domesticated as widespread crop (Østerberg et al. 2017).

Crop improvement can be achieved faster and more efficiently through knowledge of the genetic diversity that exists in the crop for selection of elite genotypes by combining genetically diverse germplasm in breeding programs and for systematic conservation of germplasm to avoid genetic erosion in gene pools (Fu et al. 2013). Currently, no published primer sequences are available for any DNA marker for *C. hindsii*. Although polymorphisms were found between populations and geographic locations of these two species in the current study using RAPD and RAMP markers, these universal sequences might not be specific enough for marker-assisted breeding. Next-generation sequencing platforms are efficient tools for identifying many DNA markers at a lower cost compared to previous sequencing methods (Ekblom & Galindo 2010). Mining NGS data for DNA marker development has been used in a variety of crops (Yang et al. 2012; Gimode et al. 2016; Ronoh et al. 2018). At present, NGS applications in ecology have greatly increased the contribution that molecular tools can make to ecological restoration and enable progress in restoration genetics and rescue beyond what is possible with current methods. The benefits, addressed in small detail by NGS, are largely due to an enhanced capacity to develop molecular markers, a significantly enhanced sampling of genomes, a greater ability to assess adaptive variation, and a new capacity to characterise species composition and/or functioning via meta-barcoding or meta-genomics as mass throughput biomonitoring techniques (Stapley et al. 2010; De Filippis 2018)

7.5.3 Cultivation and sustainable product production

C. hindsii has been used by local minority groups in Vietnam for generations because it is believed to have efficacy in the treatment of several diseases related to ulcers, tumours, and inflammation diseases (Hu et al. 2013, 2014; Ly, Shimoyamada & Yamauchi 2006; Yao-Haur Kuo, Chen & Kuo 1995). The results from this study also provided further validation to the accumulation and biosynthesis of flavonoids in leaf tissues. Several proteins identified from *C. hindsii* in the current study have important implications in human disease treatment, particularly for cancer therapy (Table 7.1).

Overall, these results indicate that *C. hindsii* can be a promising candidate for medicinal plant crop development, which potential to providing a ‘natural factory’ to produce such bioactive compounds. In fact, *C. hindsii* cultivations have been initiated in some parts of Vietnam with small-scale plantations. In a recent project, Loan (2015) estimated the revenue of AUD 2,000 per year/hectare just from raw leaf material; while the whole plant

can be used for medicinal purpose, the actual value might be higher, especially in the following years when higher productivity can be achieved. Increasing added value to medicinal plant crops by controlling regulatory factors such as plant growth regulators (PGRs), elicitors, and abiotic stresses may require a longer time-frame, and the productivity is unlikely to meet the emerging demand of the globe. To supplement the conventional systems for pharmaceutical products, molecular pharming technology can be developed to produce large quantities of pharmaceutical substances. Producing these bioactive compounds in plants has been considered easy and efficient compared to animal or microbial cell cultures, as these latter systems are associated with excessive costs associated with maintenance, safety, storage and transport (Stoger et al. 2014).

Cultivation can produce a more reliable supply of raw material and the price of the material will be more stable than with wild collected material purchased. For successful nursery propagation of *C. hindsii*, high-quality (elite plants) raw material harvested from well-designated cultivations (for example, developed from BL population from both wild and existing population) should be initiated using low input cultivation methods to compete in the local markets, and possibly in internationally, and extended investment should proceed to gain maximum added value.

Intercropping systems may be another method for small-hold farming systems. The rationale is that interspecies competition and complementation between intercrops may lead to changes in plants accumulation of mineral and secondary metabolites, and hence alter pharmaceutical and nutritional quality. Ngwene et al. (2017) revealed that some health-related phytochemicals could be modified by intercropping and late-season drought (stress). This cultivated technique increased total glucosinolate content in *Brassica carinata* intercropped with *Solanum scabrum* under low irrigation while maintaining biomass production and the content of other health-related minerals in both species. However, for effective intercropping with *C. hindsii*, field validation of phytochemical accumulation is necessary before a definite recommendation can be made to stakeholders. The potential candidates for *C. hindsii* intercrops maybe legumes, as these species have the closest phylogeny with Celastraceae, and more importantly they are good crops by their capability of fixing atmosphere nitrogen and replenishing soil fertility, which are important for soil rehabilitation (Shen et al. 2013). This system might become more attractive to indigenous people as it provides both dietary nutrients and pharmaceutical benefit for consumers.

For large-scale cultivation of *C. hindsii*, several agricultural methods of cultivation should be considered. Tissue culture and somatic embryogenesis may be applicable in selecting 'elite plants' for field nurseries, and these types of facilities must be developed (Jain & Gupta 2013; De Filippis 2014). Synthetic seed production is another useful approach where seeds can be stored and are viable for lengthy periods. However, these methods require infrastructure and initial investment and maintenance of structure for long-term effectiveness. Also, long-term monitoring and effects of seed stability and the genetic stability of elite plants may be required as the sequential steps of domestication and large-scale cultivation. In this period of genetic advancements, both traditional markers and

next-generation sequencing (NGS) markers are indispensable, however funding for further research is urgent, and must be appreciated.

Recently, molecular pharming has emerged as a contemporary technology for the manufacture of pharmaceuticals products, particularly glycoproteins, which cannot be produced efficiently in microbes or mammalian cells. Prokaryotic expression systems are generally unsuitable for therapeutic proteins, particularly monoclonal antibodies in cancer treatment, due to the lack of some post-translational modification (e.g. N-linked glycosylation, which is often required for therapeutic proteins to function properly) and inefficiency of others (e.g. disulphide bond formation, which is necessary for proteins to fold correctly). On the other hand, expression platforms based on mammalian cells may be incompatible with the production of small molecule anticancer drugs and proteins as these systems are intentionally designed to be highly toxic towards mammalian cell, e.g. by impairing cell division (Buyel 2018).

Progress in plant engineering and synthetic biology has significantly improved the awareness of using plant as production hosts leading to great efforts in the implementation and enhancement of pharmaceutical production in both *in vivo* and *in vitro* production (Webster et al. 2017). Initially, molecular pharming was highly appreciated as producing valuable recombinant proteins at a fraction of the cost of conventional systems but now remain particularly suitable for the manufacture of certain niche products (Dirisala et al. 2017). Examples include the enzyme taliglucerase alfa, a recombinant form of human glucocerebrosidase developed for the treatment of the lysosomal storage disorder Gaucher's disease; insulin produced in safflower; and an HIV-neutralising monoclonal antibody produced in tobacco (Stoger et al. 2014). Moreover, heterologous *in planta* production appears to be more cost-effective and environmentally friendly than other current biotechnological platforms (Faye & Gomord 2010; Khatri, Saini & Chhillar 2017). Since *C. hindsii* appears to possess bioactive glycoproteins, molecular pharming should be considered carefully in near- and mid-term sense as a viable option.

7.6 Conclusions and future research perspective

7.6.1 Genomic study

RAPD-PCR and RAMP-PCR were used to study the genetic variation and genetic structure of the critically endangered *C. fugax* and medicinal plant *C. hindsii*. The results suggested that both RAPD and RAMP markers are useful and informative. The ability of the markers to distinguish between populations of these two species indicated that RAMP-PCR has a better potential for use in conservation genetics, and further molecular developments would aid in this process; e.g. development of simple sequence repeats (SSRs) primers more specific for each species using next-generation sequencing technology, and single nucleotide polymorphisms (SNPs). This study is the first report providing basic genetic diversity information on these two species distributed in both natural and cultivated habitats. The results of the current study are consistent with several other studies, which use RAPD and RAMP data to generate distribution patterns and

genetic information, and suggest that molecular markers could be successfully applied for detecting genetic variability in populations of forest and domesticated species. The results could be part of a helpful 'tool kit' in conservation genetics of endangered plants and establish priorities for conservation policies and management of plant resources.

7.6.2. Metabolic study

Metabolites profiling was the method of choice to explore a selected portion of metabolome differences between NL and BL of *C. hindsii*, through spectrophotometric measurements and showed that amounts of α-tocopherol, flavonoid, phospholipid and possibly glutathione in BL leaf tissues of *C. hinsdii* was higher. Confirmation of these higher metabolite levels would benefit this study by using alternate methods (e.g. high-pressure liquid chromatography-HPLC) which could detect lower amounts of compounds. These important antioxidative metabolites play an important role in plant physiological processes to protect plants but are also widely used in prevention and treatment of various human diseases. Proline and MDA were increased in NL of *C. hindsii*, indicating that this phenotype was under stress (general or oxidative), or differences were just due to genetic variation (high genetic differentiation was revealed by genetic diversity analysis). These environmental and genetic variations may also be the sources of protein content and protease differences, in which higher protein content in BL was correlated to lower protease activity (acid protease). The results reflected that BL contains more stable and higher amount of proteins and metabolites, which maybe have pharmaceutical implications.

7.6.3 Proteomic study

Differentially expressed proteins were analysed between NL (low flavonoid content plant) and BL (high flavonoids content plant) using 2D-PAGE, coupled with mass spectrometry to generate MS/MS data for database searching. BL exhibited more diversity of proteins, and some of the upregulated proteins were found to be associated with flavonoid biosynthesis, stress response and were related to human diseases. Leucoanthocyanidin reductase, germin-like protein, acidic endochitinase, disease resistance proteins, Kunitz family trypsin and protease inhibitor, auxin-binding protein, serine/threonine-protein kinase, and protein IQ-DOMAIN, just to name a few (Table 7.1). These proteins with antimicrobial, antifungal and antiviral properties, could account for the better growth of BL. Proteins, such as dicer, β-1,3-galactosyltransferase, Kunitz-trypsin type proteinase inhibitor, staphylococcal nuclease domain-containing protein, cullin-4 may also play important roles in plant growth and development, but their more important roles are in pharmacological properties and therapeutic implications in human disease treatment, particularly cancer therapy. These proteins may be considered as important biomarkers for breeding selection and biotechnological development to define targets for medicinal plant cultivation, domestication, and molecular pharming. However, in further proteomics research, more sample replication over more plant genotypes is required to validate these conclusions.

7.6.4 Management and sustainability

Ex situ conservation and cultivation are not always practiced, therefore, the most effective way to achieve good management of resources is to educate and practice sustainable management of resources. The results from this study suggest that small population of *C. fugax* and possibly also *C. hindsii* in Vietnam are well worth conserving as part of an overall strategy to maintain genetic diversity. Management to ensure better conservation of both species, however, must be more pro-active than at present. Firstly, the significance of the study should be presented to the managers of these resources. Secondly, both *ex situ* and *in situ* conservation should be implemented in designated areas, such as national park buffer zones and protected sites, of the remaining fragmented populations. Thirdly, shallow soils of the area should be preserved as much as possible. The fourth step would be to educate the indigenous population not to collect material from the wild and develop a plan to invest in methods of propagation and farm these valuable species in domestic situations. Finally, the conservation of these endangered plants would benefit by interconnecting the present fragmented populations via corridors to aid continuing genetic exchange (Siol et al. 2010).

A comprehensive plan for management of any plant is likely to cover from one to a few of these points of views. However, this study is an attempt to focus on genetic diversity and, specifically, on the potential of new breeding and genetic research to overcome some of the concerns and constraints of rare, endangered and medicinal plants. Management of plant resources should take a temporal and spatial consideration, as plant species may suddenly become popular and have the focus of research in an area, while may be poorly studied and managed in another region. Therefore, time management is important to avoid confusion that may cause even more problems, misunderstanding and lead to poor management.

For any ethnic groups engaged in agriculture, better agricultural production is a core part of the ethnic culture, and traditional agriculture is an integral part of the culture of ethnic minorities. Therefore, *in situ* conservation and sustainability, including techniques developed for propagation, cultivation, on-farm and off-farm management, new variety breeding, and scientific studies are essential for agrobiodiversity to be maintained (Schou et al. 2017; Wayne & Bradley 2016). In many mountain/forest areas, local food and medicinal plant security are mostly dependent on conservation of agricultural biological and plant resources, so *in situ* conservation measures by local farmers can be very important. We can recommend the following conservation strategies:

(a) Promote consensus among different stakeholders on the value of agrobiodiversity and conservation.
(b) Encourage households to conserve traditional varieties in permanent plots.
(c) Convene seed exchange facilities among farmers as a method to maintain crop diversity.
(d) Make a visual documentary of the indigenous knowledge related to cultivation by recording the status, distribution, use and cultural management for policymakers.

(e) Provide traditional agricultural products to tourists to promote agrobiodiversity conservation with assistance from local governments.
(f) *Ex situ* conservation and agrobiodiversity may be appropriate for rare and endangered medicinal plants, but also provides sources of plant material for research if needed.

7.6.5 Bioprocessing and molecular pharming

C. hindsii may be a repository of unforeseen potentialities and should be studied and preserved for the benefit of present and future generations, as they are crucial to support human wellbeing by also contributing in the increase of income of the rural populations. While the production of various metabolites such as artemisinin, taxol, etc. can be extremely complicated as plant organs at different development stages produce metabolites with low yields (Chemler & Koffas 2008), industrial large-scale bioprocessing (from suspension culture, organ culture, hairy root culture) represents as a substitute for the commercial production for several phytochemicals. Perhaps the continuous production of valuable plant metabolites and proteins in suspended bioreactor cultures and *in situ* extraction to avoid cell toxicity is not too far in the future. This will prevent some regulatory restrictions placed on field plants and allow for the stable continuous production of drugs. However, there are some limitations in this approach, including cost, scalability, safety and compound authenticity. These limitations have prompted research into alternative methods such as plant-based molecular pharming approaches with lesser costs and toxicity (Buyel 2018).

Current advances in molecular pharming drive a more diverse research and development approach to improving *in planta* production of recombinant proteins (Khatri et al. 2017). In this regard, plant-based expression has emerged as an accepted alternative to conventional expression platforms due to economic feasibility, rapid scalability, higher stability of recombinant proteins, and safety thanks to a lack of harmful substances, which may result from human/animal pathogens, and a capability of producing proteins with undesirable secondary modifications (Dirisala et al. 2017). Advances in multigene transformation and transcription factors, along with targeting of cellular compartment techniques may enable elevated production levels in future engineered plants; bringing us closer to industrial scale plant factories for plant-derived bioactive compound production (Khatri, Saini & Chhillar 2017; Buyel 2018). For example, Fuentes et al. (2016) provided an efficient strategy for engineering a complex biochemical pathway and optimising artemisinin production in crop tobacco. However, in medicinal plants, a major restriction of biotechnology is due to less knowledge about molecular tool and protocols for transgenic approaches and marker-assisted selection. Fortunately, the presence of large EST libraries and standardised protocols for most crop species transgenics, and promising NGS have made significant advances in characterising molecular pathways of bioactive compounds (Ciura et al. 2017; De Filippis 2017)

7.6.6 Next-generation sequencing (NGS)

Future improvement in NGS approaches will be the use of tagged samples (e.g. meta-barcodes), wherein multiple individuals/population can be genotyped simultaneously. These approaches are predicted to change the way we investigate plant breeding, phylogeography, demography, and conservation genetics, by massively increasing the number of loci studied (Grover et al. 2012). This improvement will require the development of new software tools to make the analyses feasible from a computational point of view. Realistically, genomics will be applied to plant breeding, biogeography, ethnobotany, and ecology when it becomes easy and cost-effective to do so (Ekblom & Galindo 2011, Yang et al. 2012; De Filippis 2018; Ronoh et al. 2018).

As the amount and quality of sequence information generated per run keep increasing, which allows even higher-level analysis and lower costs per samples, it can be anticipated that the high density of markers from NGS will be extensively applied to all plant methods, old and new (De Filippis 2018). NGS provide enormous new potential for delineating locally adapted source populations for restoration and propagation through investigating natural selections in plants using population genomic approach (Ekblom & Galindo 2011; De Filippis 2017, 2018), especially for non-model plants like *C. fugax* and *C. hindsii*. However, a major challenge for models used to infer selection comes from confounding demographic variation (e.g. changes in population size and sub-division), which is common in plant populations. Although there is great potential in NGS technology, ultimately it is the science that needs to direct the research, and it will continue to do so in the foreseeable future. To answer many research questions in ecological genetics, current tools such as RAMP-PCR, AFLP and microsatellites and SNPs will continue to make significant cost-effective contributions, as will traditional field, nursery and cultivation trials.

7.6.7 Future research not often contemplated

Based on the discussion above, we recommend that other research should be attempted:

- Further genomic, transcriptomic, and proteomic studies on fresh explants material (leaf, stem, and root) to determine and validate the expression of proteins and metabolites, since proteins at least were significantly degraded in air dried tissue.
- Determine a better understanding of the biosynthesises of other potential secondary metabolites, proteins, and bioactive compounds, e.g. elite genotypes and miRNA.
- Safety studies need to be carried out before the use of secondary metabolites and proteins, and safety assessed in extracts for pharmaceutical applications, i.e. purity.
- More studies to determine the mechanisms and possible control, and enhancement of pharmacological activities of possible bioactive compounds and extracts.
- More investigations on the isolation methods and identification of bioactive constituents and their interactions with traditional medicine and new/old drugs.

- Clinical trials, both on experimental animals and humans will need to be carried out to evaluate the efficacy of all bioactive compounds and extracts for different diseases, as most are likely to be toxic and to determine the safe and effective dosage as medicines.

REFERENCES

Abbott, R., 2014. *Documenting Traditional Medical Knowledge*. World Intellectual Property Organization.

Abdallah, C. et al., 2012. Gel-based and gel-free quantitative proteomics approaches at a glance. *International Journal of Plant Genomics*, 2012, pp.1-17.

Abreu, I.A. et al., 2013. Coping with abiotic stress: Proteome changes for crop improvement. *Journal of Proteomics*, 93, pp.145–168.

Abril, N. et al., 2011. Proteomics research on forest trees, the most recalcitrant and orphan plant species. *Phytochemistry*, 72(10), pp.1219–1242.

Adhikari, S., Bandopadhyay, T.K. & Ghosh, P.D., 2013. Assessment of genetic diversity of certain Indian elite clones of *Cymbopogon* species through RAPD analysis. *Indian Journal of Biotechnology*, 12(1), pp.109–114.

Afendi, F.M. et al., 2012. KNApSAcK family databases: Integrated metabolite-plant species databases for multifaceted plant research. *Plant and Cell Physiology*, 53(2), pp.1–12.

Agati, G. et al., 2013. Functional roles of flavonoids in photoprotection: New evidence, lessons from the past. *Plant Physiology and Biochemistry*, 72, pp.35–45.

Aghaei, K. & Komatsu, S., 2013. Crop and medicinal plants proteomics in response to salt stress. *Frontiers in Plant Science*, 4(8), pp.1-9.

Agrawal, G.K. et al., 2013. A decade of plant proteomics and mass spectrometry: translation of technical advancements to food security and safety issues. *Mass Spectrometry Reviews*, 32(5), pp.335–65.

Aguilar-Støen, M. & Moe, S.R., 2007. Medicinal plant conservation and management: Distribution of wild and cultivated species in eight countries. *Biodiversity and Conservation*, 16(6), pp.1973–1981.

Ahmad, A. & Ahmad, R., 2014. A New method for proteome screening with two-dimensional urea-SDS polyacrylamide gel electrophoresis. *Organelles Proteomics*, 1(1), pp.28–38.

Ahmad, P. et al., 2016. Role of proteomics in crop stress tolerance. *Frontiers in Plant Science*, 7(September), pp.1–14.

Aksakal, O. et al., 2017. Effect of proline on biochemical and molecular mechanisms in lettuce (*Lactuca sativa* L.) exposed to UV-B radiation. *Photochemical & Photobiological Sciences*, 16, pp.246–254.

Al-Rawashdeh, I.M., 2011. Molecular taxonomy among *Mentha spicata*, *Mentha longifolia* and *Ziziphora tenuior* populations using the RAPD technique. *Jordan Journal of Biological Sciences*, 4(2), pp.63–70.

Alam, I. et al., 2013. An improved plant leaf protein extraction method for high-resolution two-dimensional polyacrylamide gel electrophoresis and comparative proteomics. *Biotechnic & Histochemistry,* 88(2), pp.61–75.

Alasalvar, C. & Bolling, B.W., 2015. Review of nut phytochemicals, fat-soluble bioactives, antioxidant components and health effects. *British Journal of Nutrition*, 113, pp.68–78.

Ali, G. & Ghasemzadeh, N., 2011. Flavonoids and phenolic acids: Role and biochemical activity in plants and human. *Journal of Medicinal Plants Research*, 5(32), pp.6697–6703.

Ali, Z. et al., 2013. Intra-specific genetic diversity and chemical profiling of different accessions of *Clitoria ternatea* L. *Industrial Crops and Products*, 43(1), pp.768–

773.

Alia, E.A. & Gahiza, I.A., 2007. Accumulation of amino acids in *Anabaena oryzae* in response to sodium chloride salinity. *Journal of Applied Sciences Research*, 3, pp.263–266.

Aljanabi, S.M. & Martinez, I., 1997. Universal and rapid salt-extraction of high-quality genomic DNA for PCR-based techniques. *Nucleic Acids Research*, 25(22), pp.4692–3.

Allendorf, F.W., 2017. Genetics and the conservation of natural populations: allozymes to genomes. *Molecular Ecology*, 26(2), pp.420–430.

Allendorf, F.W., Hohenlohe, P. a & Luikart, G., 2010. Genomics and the future of conservation genetics. *Nature Review Genetics*, 11(10), pp.697–709.

Alnahdi, H.S., 2012. Isolation and screening of extracellular proteases produced by new isolated *Bacillus sp* environmental samples and screened their capability of protease. *Journal of Applied Pharmaceutical Science*, 2(9), pp.71–74.

Alva, V. et al., 2016. The MPI bioinformatics toolkit as an integrative platform for advanced protein sequence and structure analysis. *Nucleic Acids Research*, 44(W1), pp.W410–W415.

Amado, M. et al., 1998. A family of human β4-galactosyltransferases. *Journal of Biological Chemistry*, 273(21), pp.12770–12778.

Amalesh, S., Das, G. & Das, K.S., 2011. Roles of flavonoids in plants. *Int J Pharm Sci Tech*, 6(1), pp.12–35.

Amjad, M.W. et al., 2017. Recent advances in the design, development, and targeting mechanisms of polymeric micelles for delivery of siRNA in cancer therapy. *Progress in Polymer Science*, 64, pp.154–181.

Amudha, P., Jayakumar, M. & Kulandaivelu, G., 2005. Impacts of ambient solar UV (280-400nm) radiation on three tropical legumes. *Journal of Plant Biology*, 48(3), pp.284–291.

Anca, I.A. et al., 2014. Different tobacco retrotransposons are specifically modulated by the elicitor cryptogein and reactive oxygen species. *Journal of Plant Physiology*, 171(16), pp.1533–1540.

Andersen, J.R. & Lübberstedt, T., 2003. Functional markers in plants. *Trends in Plant Science*, 8(11), pp.554–560.

Angel, T.E. et al., 2012. Mass spectrometry-based proteomics: existing capabilities and future directions. *Chemical Society Reviews*, 41(10), pp.3912–28.

Anguraj Vadivel, A.K., 2015. Gel-based proteomics in plants: time to move on from the tradition. *Frontiers in Plant Science*, 6(May), pp.2–5.

Anno, A.D. et al., 1998. Nucleic acid (DNA, RNA) quantification and RNA, DNA ratio determination in marine sediments: Comparison of high-performance liquid chromatography: methods and estimation of detrital DNA nucleic acid (DNA, RNA). *Applied and Environmental Microbiology*, 64(9), pp.3238-3245.

Anon, 1991. *Investment plan for Ba Vi National Park*, Hanoi.

Ao, Y. et al., 2012. *Celastrus yuloensis* (Celastraceae), a new species from China. *Annales Botanici Fennici*, 49(4), pp.267–270.

Aranjuelo, I. et al., 2011. Plant physiology and proteomics reveals the leaf response to drought in alfalfa (*Medicago sativa* L.). *Journal of Experimental Botany*, 62(1), pp.111–123.

Arbona, V. et al., 2013. Metabolomics as a tool to investigate abiotic stress tolerance in plants. *International Journal of Molecular Sciences*, 14(3), pp.4885–4911.

Arcade, a. et al., 2000. Application of AFLP, RAPD and ISSR markers to genetic mapping of European and Japanese larch. *Theoretical and Applied Genetics*, 100(2), pp.299–307.

ArchMiller, A.A. et al., 2015. Formalizing the definition of meta-analysis in molecular ecology. *Molecular Ecology*, 24(16), pp.4042–4051.

Armengaud, J. et al., 2014. Non-model organisms, a species endangered by proteogenomics. *Journal of Proteomics*, 105, pp.5–18.

Ashraf, M. & Foolad, M.R., 2007. Roles of glycine betaine and proline in improving plant abiotic stress resistance. *Environmental and Experimental Botany*, 59(2), pp.206–216.

Asirvatham, V.S., Watson, B.S. & Sumner, L.W., 2002. Analytical and biological variances associated with proteomic studies of Medicago truncatula by two-dimensional polyacrylamide gel electrophoresis. *Proteomics*, 2(8), pp.960–968.

Atanasov, A.G. et al., 2015. Discovery and resupply of pharmacologically active plant-derived natural products: A review. *Biotechnology Advances*, 33(8), pp.1582–1614.

Avila-Treviño, J.A. et al., 2017. In vitro propagation from bud and apex explants of Moringa oleifera and evaluation of the genetic stability with RAMP marker. *South African Journal of Botany*, 108, pp.149–156.

Ayala, A., Munoz, M.F. & Arguelles, S., 2014. Lipid peroxidation: Production, metabolism, and signalling mechanisms of malondialdehyde and 4-hydroxy-2-nonenal. *Oxidative Medicine and Cellular Longevity*, 2014, pp.1-31.

Azzi, A., 2007. Molecular mechanism of α -tocopherol action. *Free Radical Biology and Medicine*, 43, pp.16–21.

Azzi, A. & Stocker, A., 2000. Vitamin E: non-antioxidant roles. *Progress in Lipid Research*, 39, pp.231–255.

Bai, C. et al., 1996. SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-Box. *Cell*, 86(2), pp.263–274.

Bai, M. et al., 2012. Genome-wide identification of Dicer-like, Argonaute and RNA-dependent RNA polymerase gene families and their expression analyses in response to viral infection and abiotic stresses in *Solanum lycopersicum*. *Gene*, 501(1), pp.52–62.

Bamigboye, S.O., Tshisikhawe, P.M. & Taylor, P.J., 2017. Detecting threats to *Encephalartos transvenosus* (Limpopo cycad) in Limpopo province, South Africa through indigenous knowledge. *Indian Journal of Traditional Knowledge*, 16(2), pp.251–255.

Ban, N. et al., 2007. *Vietnam Red Book Part II. Plants*, Natural Sciences and Technology Publishers, Hanoi: [In Vietnamese].

Ban, N.T., 2003. *Catalogue of Vietnam's plant species* Volume II., Hanoi: Agriculture Publishing House: [In Vietnamese].

Banothu, V., Adepally, U.M.A. & Lingam, J., 2017. In vitro total phenolics, flavonoids contents, antioxidant and antimicrobial activities of various solvent extracts from the medicinal plant P*hysalis minima* Linn. *International Journal of Pharmaceutical Sciences*, 9(3), pp.192-198.

Bantscheff, M. & Lemeer, S., 2012. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Analytical and Bioanalytical Chemistry*, 404(4), pp.939–965.

Bantscheff, M. & Schirle, M., 2007. Quantitative mass spectrometry in proteomics: a critical review. *Analytical and Bioanalytical Chemistry*, 389(4), pp.1017–1031.

Bao, F. et al., 2017. Overexpression of *Prunus mume* dehydrin genes in tobacco enhances tolerance to cold and drought. *Frontiers in Plant Science*, 8(February), pp.1–12.

Barata, A.M. et al., 2016. Conservation and sustainable uses of medicinal and aromatic plants genetic resources on the worldwide for human welfare. *Industrial Crops and Products*, 88, pp.8–11.

Barcelo, J. & Poschenrieder, C., 2002. Fast root growth responses, root exudates, and internal detoxification as clues to the mechanisms of aluminium toxicity and resistance: a review. *Environmental and Experimental Botany*, 48, pp.75–92.

Bardakci, F., 2001. Random Amplified Polymorphic DNA (RAPD) Markers. *Turk J Biol*, 25, pp.185–196.

Barkan, A. & Small, I., 2014. Pentatricopeptide repeat proteins in plants. *Annual Review of Plant Biology*, 65(1), pp.415–442.

Baruah, J. et al., 2017. Genetic diversity study amongst *Cymbopogon* species from NE-India using RAPD and ISSR markers. *Industrial Crops and Products*, 95, pp.235–243.

Bastos, P. et al., 2017. Unveiling antimicrobial peptide–generating human proteases using PROTEASIX. *Journal of Proteomics*, 171, pp.53-62.

Basu, D. et al., 2015. Two hydroxyproline galactosyltransferases, GALT5 and GALT2, function in growth and development in A*rabidopsis*. *PLoS ONE*, 10(5), pp.1–36.

Bates, L.S., Waldren, R.P. & Teare, I.D., 1973. Rapid determination of free proline for water-stress studies. *Plant and Soil*, 39(1), pp.205–207.

Beach, J.R. & Hammer, J.A., 2015. Myosin II isoform co-assembly and differential regulation in mammalian systems. *Experimental Cell Research*, 334(1), pp.2–9.

Beattie, J.R. et al., 2017. The use of raman microscopy to determine and localise vitamin E in biological samples. *The FASEB Journal*, 21(3), pp.766–776.

Bennetzen, J.L., 2000. Comparative sequence analysis of plant nuclear genomes microcolinearity and its many exceptions. *The Plant Cell*, 12(7), pp.1021–9.

Bento, J.F., Noleto, G.R. & De Oliveira Petkowicz, C.L., 2014. Isolation of an arabinogalactan from *Endopleura uchi* bark decoction and its effect on HeLa cells. *Carbohydrate Polymers*, 101(1), pp.871–877.

Betapudi, V., Licate, L.S. & Egelhoff, T.T., 2006. Distinct roles of nonmuscle myosin II isoforms in the regulation of MDA-MB-231 breast cancer cell spreading and migration. *Cancer Research*, 66(9), pp.4725–4733.

Bhagat, J., Ingole, B.S. & Singh, N., 2016. Glutathione S-transferase, catalase, superoxide dismutase, glutathione peroxidase, and lipid peroxidation as biomarkers of oxidative stress in snails: A review. *ISJ*, 13, pp.336–349.

Bhattacharjee, R. et al., 2004. An improved semiautomated rapid method of extracting genomic DNA for molecular marker analysis in Cocoa, *Theobroma cacao* L. *Plant Molecular Biology Reporter*, 22, p.435a–435h.

Bhattacharyya, D. et al., 2012. Proteins differentially expressed in elicited cell suspension culture of Podophyllum hexandrum with enhanced podophyllotoxin content. *Proteome Science*, 10(1), p.34.

Bian, F. et al., 2009. Proteomic analysis of somatic embryogenesis in *Cyclamen persicum* Mill. *Plant Molecular Biology Reporter*, 28(1), pp.22–31.

Bian, S. & Jiang, Y., 2009. Reactive oxygen species, antioxidant enzyme activities and gene expression patterns in leaves and roots of Kentucky bluegrass in response to drought stress and recovery. *Scientia Horticulturae*, 120(2), pp.264–270.

Bienert, M.D. et al., 2012. NtSCP1 from Tobacco is an extracellular serine

carboxypeptidase III that has an impact on cell elongation. *Plant Physiology*, 158(3), pp.1220–1229.

Bijina, B. et al., 2011. Protease inhibitor from *Moringa oleifera* leaves: Isolation, purification, and characterisation. *Process Biochemistry*, 46(12), pp.2291–2300.

Birt, T.P. & Baker, A.J., 2000. Polymerase chain reaction. In A. J. Baker, ed. *Molecular Methods in Ecology*. Oxford: Blackwell Science.

Bishoyi, A.K. et al., 2016. Varietal discrimination and genetic variability analysis of *Cymbopogon* using RAPD and ISSR markers analysis. *Applied Biochemistry and Biotechnology*, 179(4), pp.659–670.

Blair, I.A., 2008. DNA adducts with lipid peroxidation products. *Journal of Biological Chemistry*, 283(23), pp.15545–15549.

Blanco, M.A. et al., 2011. Identification of staphylococcal nuclease domain-containing 1 (SND1) as a metadherin-interacting protein with metastasis-promoting functions. *Journal of Biological Chemistry*, 286(22), pp.19982–19992.

Bodzon-Kulakowska, A. et al., 2007. Methods for samples preparation in proteomic research. *Journal of Chromatography*, 849(1–2), pp.1–31.

Bota, C. & Deliu, C., 2015. Effect of plant growth regulators on the production of flavonoids by cell suspension cultures of *Digitalis lanata*. *Farmacia*, 63(5), pp.716–719.

Boughton, B.A. et al., 2016. Mass spectrometry imaging for plant biology: a review. *Phytochemistry Reviews*, 15(3), pp.445–488.

Bourgeois, M. et al., 2009. Dissecting the proteome of mature pea seeds reveals the phenotypic plasticity of seed protein composition. *Proteomics*, 9(2), pp.254–271.

Boutet, E. et al., 2016. UniProtKB/Swiss-Prot, the manually annotated section of the UniProt knowledgeBase: How to use the entry view. In D. Edwards, ed. *Plant Bioinformatics: Methods and Protocols*. New York, NY: Springer New York, pp. 23–54.

Boye, J.I. et al., 2010. Comparison of the functional properties of pea, chickpea and lentil protein concentrates processed using ultrafiltration and isoelectric precipitation techniques. *Food Research International*, 43(2), pp.537–546.

Bracci, T. et al., 2011. Molecular studies in olive (*Olea europaea* L.): overview on DNA markers applications and recent advances in genome analysis. *Plant Cell Reports*, 30(4), pp.449–62.

Breton, C. et al., 2006. Structures and mechanisms of glycosyltransferases. *Glycobiology*, 16(2), pp.29–37.

Brooks, A., 2006. Enhancing the effectiveness of projects on Cat Ba Island-an evaluation of ten years of international support. *IUCN Vietnam Country Office: Hanoi, Vietnam*.

Brown, A.H., 1978. Isozymes, plant population genetic structure and genetic conservation. *Theoretical and Applied Genetics*, 52, pp.145–57.

Brown, A.H.D., 1979. Enzyme polymorphism in plant populations. *Theoretical Population Biology*, 15(1), pp.1–42.

Brunetti, C. et al., 2013. Flavonoids as antioxidants and developmental regulators : Relative significance in plants and humans. *International Journal of Molecular Sciences*, 14, pp.3540–3555.

Bryant, L., Patole, C. & Cramer, R., 2016. Proteomic analysis of the medicinal plant *Artemisia annua*: Data from leaf and trichome extracts. *Data in Brief*, 7, pp.325–331.

Buchner, O. et al., 2017. Drought affects the heat-hardening capacity of alpine plants as indicated by changes in xanthophyll cycle pigments, singlet oxygen scavenging, α-tocopherol and plant hormones. *Environmental and Experimental Botany*, 133, pp.159–175.

Bui, D.T. et al., 2016. Tropical forest fire susceptibility mapping at the Cat Ba National Park area, Hai Phong city, Vietnam. *Remote Sensing*, 8(347), pp.1–15.

Burbulis, N. et al., 2011. Changes in proline and soluble sugars content during cold acclimation of winter rapeseed shoots *in vitro*. *Journal of Food, Agriculture and Environment*, 9(2), pp.371–374.

Burritt, D.J., 2012. Proline and the cryopreservation of plant tissues: Functions and practical applications. In Katkov, I., ed. *Current Frontiers in Cryopreservation*. In Tech, pp.415-427.

Bürstenbinder, K. et al., 2017. The IQD family of calmodulin-binding proteins links calcium signalling to microtubules, membrane subdomains, and the nucleus. *Plant Physiology*, 173(3), pp.1692–1708.

Buse, J. et al., 2015. Conservation of threatened habitat types under future climate change – Lessons from plant-distribution models and current extinction trends in southern Germany. *Journal for Nature Conservation*, 27, pp.18–25.

Buts, K. et al., 2014. Improving the identification rate of data independent label-free quantitative proteomics experiments on non-model crops: A case study on apple fruit. *Journal of Proteomics*, 105, pp.31–45.

Buyel, J.F., 2018. Plants as sources of natural and recombinant anti-cancer agents. *Biotechnology Advances*, 36(2), pp.506–520.

Borner, A., Chebotar, S. & Korzun, V., 2000. Molecular characterisation of the genetic integrity of wheat (*Triticum aestivum* L.) germplasm after long-term maintenance. *Theoretical and Applied Genetics*, 100(3–4), pp.494–497.

Calonje, M. et al., 2011. Propagation of cycad collections from seed: applied reproductive biology for conservation. *The Journal of Botanic Garden Horticulture*, 9, pp.77-96.

Canut, H., Albenne, C. & Jamet, E., 2016. Post-translational modifications of plant cell wall proteins and peptides: A survey from a proteomics point of view. *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1864(8), pp.983–990.

Cardinale, B.J. et al., 2012. Corrigendum: Biodiversity loss and its impact on humanity. *Nature*, 489(7415), pp.326–326.

Carocho, M. & Ferreira, I.C.F.R., 2013. A review on antioxidants, prooxidants and related controversy: natural and synthetic compounds, screening and analysis methodologies and future perspectives. *Food and Chemical Toxicology*, 51, pp.15–25.

Carpentier, S.C. et al., 2005. Preparation of protein extracts from recalcitrant plant tissues: An evaluation of different methods for two-dimensional gel electrophoresis analysis. *Proteomics*, 5(10), pp.2497–2507.

Carrington, J.C., 2003. Role of microRNAs in plant and animal development. *Science*, 301(5631), pp.336–338.

Carstens, M., Vivier, M.a. & Pretorius, I.S., 2003. The *Saccharomyces cerevisiae* chitinase, encoded by the CTS1-2 gene, confers antifungal activity against *Botrytis cinerea* to transgenic tobacco. *Transgenic Research*, 12(4), pp.497–508.

Carvalho, K.S. et al., 2013. Diurnal changes in leaflet gas exchange, water status and antioxidant responses in Carapa guianensis plants under water-deficit conditions. *ACTA Physiologiae Plantarum*, 35(1), pp.13–21.

Catusse, J. et al., 2011. Proteomics reveals potential biomarkers of seed vigour in sugarbeet. *Proteomics*, 11(9), pp.1569–1580.

Cercós, M., Urbez, C. & Carbonell, J., 2003. A serine carboxypeptidase gene (PsCP), expressed in early steps of reproductive and vegetative development in *Pisum sativum*, is induced by gibberellins. *Plant Molecular Biology*, 51(2), pp.165–174.

Chaisurisri, K. & El-Kassaby, Y.A., 1994. Genetic diversity in a seed production population vs. natural populations of *Sitka spruce*. *Biodiversity and Conservation*, 523, pp.512–523.

Chakrabarti, M. et al., 2013. Overexpression of miR-7-1 increases efficacy of green tea polyphenols for induction of apoptosis in human malignant neuroblastoma SH-SY5Y and SK-N-DZ cells. *Neurochemical Research*, 38(2), pp.420–432.

Chalmers, K.J. et al., 1994. Evaluation of the extent of genetic variation in *Mahoganies* (Meliaceae) using RAPD markers. *Theoretical and Applied Genetics*, 89(4), pp.504–8.

Champagne, A. et al., 2012. In-depth proteome mining of cultured *Catharanthus roseus* cells identifies candidate proteins involved in the synthesis and transport of secondary metabolites. *Proteomics*, 12(23–24), pp.3536–3547.

Champagne, A. & Boutry, M., 2013. Proteomics of nonmodel plant species. *Proteomics*, 13(3–4), pp.663–673.

Chandra, S. et al., 2017. *In silico* characterisation, homology modelling of *Camellia sinensis* chitinase and its evolutionary analyses with other plant chitinases. *Proceedings of the National Academy of Sciences India Section B - Biological Sciences*, 87(3), pp.685–695.

Chase, M.W. & Hills, H.H., 1991. Silica gel: An ideal material for field preservation of leaf samples for DNA studies. *Taxon*, 40(2), pp.215–220.

Chatti, K. et al., 2007. Analysis of genetic diversity and relationships in a Tunisian fig (*Ficus carica*) germplasm collection by randomly amplified microsatellite polymorphisms. *Journal of Integrative Plant Biology*, 49(3), pp.386–391.

Chaudhry, B., Yasmeen, A. & Husnain, T., 1999. Mini-scale genomic DNA extraction from Cotton. *Plant Molecular Biology Reporter,* 17(3), pp.280-280.

Chemler, J.A. & Koffas, M.A.G., 2008. Metabolic engineering for plant natural product biosynthesis in microbes. *Current Opinion in Biotechnology*, 19(6), pp.597–605.

Chen, D. & Ronald, P.C., 1999. A rapid DNA minipreparation method suitable for AFLP and other PCR applications. *Plant Molecular Biology Reporter*, 17, pp.53–57.

Chen, J.L. et al., 2011. Artemisinin biosynthesis enhancement in transgenic *Artemisia annua* plants by downregulation of the β-caryophyllene synthase gene. *Planta Medica*, 77(15), pp.1759–1765.

Chen, Q., Zhang, M. & Shen, S., 2011. Comparison of protein extraction methods suitable for proteomics analysis in seedling roots of Jerusalem artichoke under salt (NaCl) stress. *African Journal of Biotechnology*, 10(39), pp.7650–7657.

Chen, S. & Harmon, A.C., 2006. Advances in plant proteomics. *Proteomics*, 6(20), pp.5504–5516.

Chen, S.L. et al., 2016. Conservation and sustainable use of medicinal plants: problems, progress, and prospects. *Chinese Medicine*, pp.1–10.

Chen, W.P. & Li, P.H., 2002. Membrane stabilisation by abscisic acid under cold aids proline in alleviating chilling injury in maize (*Zea mays* L.) cultured cells. *Plant, Cell and Environment*, 25(8), pp.955–962.

Chen, Y. et al., 2014. Genetic linkage map construction for white jute (*Corchorus*

capsularis L.) using SRAP, ISSR and RAPD markers. *Plant Breeding*, 133, pp.777-781.

Chen, Y. et al., 2013. The F-Box protein OsFBK12 targets OsSAMS1 for degradation and affects pleiotropic phenotypes, including leaf senescence, in rice. *Plant Physiology*, 163(4), pp.1673-1685.

Chen, Z. et al., 2015. Cullin family proteins and tumorigenesis: Genetic association and molecular mechanisms. *Journal of Cancer*, 6(3), pp.233–242.

Cheng, J.S. & Yuan, Y.J., 2006. Proteomic analysis reveals the spatial heterogeneity of immobilised Taxus cuspidata cells in support matrices. *Proteomics*, 6(7), pp.2199–2207.

Cheung, S.K.C. et al., 2016. Stage-specific embryonic antigen-3 (SSEA-3) and β3GalT5 are cancer-specific and significant markers for breast cancer stem cells. *Proceedings of the National Academy of Sciences*, 113(4), pp.960–965.

Chevin, L.M. & Hoffmann, A.A., 2017. Evolution of phenotypic plasticity in extreme environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1723).

Chin, A.R. et al., 2016. Cross-kingdom inhibition of breast cancer growth by plant miR159. *Cell Research*, 26, pp.217-228.

Chittapalo, T. & Noomhorm, A., 2009. Ultrasonic assisted alkali extraction of protein from defatted rice bran and properties of the protein concentrates. *International Journal of Food Science & Technology*, 44(9), pp.1843–1849.

Cho, Y.A. et al., 2017. Dietary flavonoids, CYP1A1 genetic variants, and the risk of colorectal cancer in a Korean population. *Scientific Reports*, 7(1), pp.1–8.

Choudhury, F.K. et al., 2016. Reactive oxygen species, abiotic stress and stress combination. *The Plant Journal*, 90(5), pp.1–12.

Cijo, V., Dellaire, G. & Rupasinghe, H.P.V., 2017. Plant flavonoids in cancer chemoprevention: role in genome stability. *The Journal of Nutritional Biochemistry*, 45, pp.1–14.

Cipollini, K. et al., 2013. Genetic structure of the endangered northeastern bulrush (*Scirpus ancistrochaetus*) in Pennsylvania, USA, using information from RAPD and SNPs. *Biochemical Genetics*, 51(9–10), pp.686–697.

Citovsky, V. & Liu, B., 2017. Myosin-driven transport network in plants is functionally robust and distinctive. *Proceedings of the National Academy of Sciences*, 114(8), pp.1756–1758.

Ciura, J. et al., 2017. Next-generation sequencing of representational difference analysis products for identification of genes involved in diosgenin biosynthesis in fenugreek (*Trigonella foenum-graecum*). *Planta*, 245(5), pp.977–991.

Clarke, K.R. & Gorley, R.N., 2001. *Primer V5: user manual/tutorial*, Primer-E Ltd. The UK.

Coletta, A. et al., 2014. Effect of viticulture practices on the concentration of polyphenolic compounds and total antioxidant capacity of Southern Italy red wines. *Food Chemistry*, 152, pp.467–74.

Collard, B.C.Y. & Mackill, D.J., 2008. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Phil Trans R Soc B*, 363(8), pp.557–572.

Colombo, C., Second, G. & Charrier, A., 2000. Diversity within American cassava germ plasm based on RAPD markers. *Genetics and Molecular Biology*, 23, pp.189–199.

Corchete, P. & Bru, R., 2013. Proteome alterations monitored by DIGE analysis in

Silybum marianum cell cultures elicited with methyl jasmonate and methyl B cyclodextrin. *Journal of Proteomics*, 85, pp.99–108.

Corle, R.T., 2016. Plant diversity in a changing world. *Plant Diversity*, 38, pp.11–18.

Del Cornò, M. et al., 2017. Linking diet to colorectal cancer: The emerging role of microRNA in the communication between plant and animal kingdoms. *Frontiers in Microbiology*, 8(APR).

Cosgrove, D.J., 2014. Re-constructing our models of cellulose and primary cell wall assembly. *Current Opinion in Plant Biology*, 22, pp.122–131.

Couch, J. a. & Fritz, P.J., 1990. Isolation of DNA from plants high in polyphenolics. *Plant Molecular Biology Reporter*, 8(1), pp.8–12.

Cousins, S.R., Williams, V.L. & Witkowski, E.T.F., 2011. Quantifying the trade in cycads (*Encephalartos Species)* in the traditional medicine markets of Johannesburg and Durban, South Africa. *Economic Botany*, 65(4), pp.356–370.

Cousins, S.R. & Witkowski, E.T.F., 2017. African cycad ecology, ethnobotany and conservation: A synthesis. *Botanical Review*, 83(2), pp.152–194.

Cramer, R., Bindschedler, L. & Agrawal, G., 2013. Plant proteomics in crop improvement. *Proteomics*, 13(12–13), pp.1771–1771.

Crifò, T. et al., 2011. Expression analysis in response to low-temperature stress in blood oranges: Implication of the flavonoid biosynthetic pathway. *Gene*, 476(1), pp.1–9.

Cui, D., Xiong, X. & Zhao, Y., 2016. Cullin-RING ligases in regulation of autophagy. *Cell Division*, 11(8), pp.1-14.

Cui, L. & Decker, E.A., 2016. Phospholipids in foods : prooxidants or antioxidants ? *J Sci Food Agric*, 96, pp.18–31.

Cui, L., Mcclements, D.J. & Decker, E.A., 2015. Impact of phosphatidylethanolamine on the antioxidant activity of alpha-tocopherol and Trolox in bulk oil. *Journal of Agricultural and Food Chemistry*, 63(12), pp.3288–3294.

Currò, S. et al., 2010. New microsatellite loci for pomegranate, *Punica granatum* (Lythraceae). *American Journal of Botany*, 97(7), pp.e58-60.

Da Silva Bezerra, C. et al., 2016. Exploiting the biological roles of the trypsin inhibitor from *Inga vera* seeds: A multifunctional Kunitz inhibitor. *Process Biochemistry*, 51(6), pp.792–803.

Dacaranhe, C.D. & Terao, J., 2001. A unique antioxidant activity of phosphatidylserine on iron-induced lipid peroxidation of phospholipid bilayers. *Lipids*, 36(10), pp.1105–1110.

Dai, Q. & Wang, H., 2006. Cullin 4 makes its mark on chromatin. *Cell Division*, 1(14), pp.1–7.

Damasco, O.P. et al., 1996. Random amplified polymorphic DNA (RAPD) detection of dwarf off-types in micro-propagated cavendish (*Musa spp.*) bananas. *Plant Cell Reports*, 16, pp.118–123.

Damerval, C. et al., 1986. Technical improvements in two-dimensional electrophoresis increase the level of genetic variation detected in wheat-seedling proteins. *Electrophoresis*, 7(1), pp.52–54.

Darwish, E. et al., 2009. Phospholipid signalling responses in salt-stressed rice leaves. *Plant and Cell Physiology*, 50(5), pp.986–997.

Das, K. & Roychoudhury, A., 2014. Reactive oxygen species (ROS) and response of antioxidants as ROS-scavengers during environmental stress in plants. *Frontiers in Environmental Science*, 2(December), pp.1–13.

Dávila, J. a et al., 1999. Comparison of RAMP and SSR markers for the study of wild barley genetic diversity. *Hereditas*, 131(1), pp.5–13.

Day, P.R., 1996. The biology of plant proteins. *Critical Reviews in Food Science and Nutrition*, 36(sup001), pp.39–47.

De Filippis, L.F., 2013. Bioinformatic tools in crop improvement. In K. R. Hakeem, P. Ahmad, & M. Ozturk, eds. *Crop improvement: New approaches and modern techniques*. Boston, MA: Springer US, pp. 49–122.

De Filippis, L.F, 2017a. Crop improvement by tissue culture. In H. KR, A. PA, & O. MA, eds. *Crop Improvement - New approaches and Modern Techniques*. Springer Business Dordrecht, pp. 49–122.

De Filippis, L.F., 2017. Plant bioinformatics: next-generation sequencing approaches. In F. Hakeem, KR, Malik, A, Ozturk, MA, Sukan, ed. *Plant Bioinformatics: Decoding the Phyta*. Dordrecht: Springer Business.

De Filippis, L.F., 2018. Underutilised and neglected crops: Next-generation sequencing approaches for crop improvement and better food security. In H. KR et al., eds. *Plant Bioinformatics: Decoding the Phyta*. Springer Business Dordrecht, pp. 1–44.

De Filippis, L.F., Hoffmann, E. & Hampp, R., 1996. Identification of somatic hybrids of tobacco-generated by electrofusion and culture of protoplasts using RAPD-PCR. *Plant Science*, 121(1), pp.39–46.

De Filippis, L.F. & Magel, E., 1998. Differences in genomic DNA extracted from bark and from wood of different zones in Robinia trees using RAPD-PCR. *Trees*, 12(6), pp.377–384.

De Filippis, L.F. & Magel, E., 2012. Identification of biochemical differences between the sapwood and transition zone in Robinia pseudoacacia L. by differential display of proteins. *Holzforschung*, 66(4), pp.543–549.

Debnath, T. et al., 2017. Role of phytochemicals in the modulation of miRNA expression in cancer. *Food & Function*, 8(10), pp.3432–3442.

De O Maia, I.R. et al., 2017. Content of total phenolic compounds, flavonoids and tannins in methanol extracts of the genus S*enna* Mill. from the northeast of Brazil and evaluation of antioxidant capacity. *Journal of Pharmacognosy and Phytochemistry*, 6(5), pp. 1321-1325.

Decker, G. et al., 2000. Characterization of proteins in latex of the opium poppy (*Papaver somniferum*) using two-dimensional gel electrophoresis and microsequencing. *Electrophoresis*, 21(16), pp.3500–16.

Degl'Innocenti, E. et al., 2005. Biochemical study of leaf browning in minimally processed leaves of lettuce (*Lactuca sativa* L. *var. acephala*). *Journal of Agricultural and Food Chemistry*, 53(26), pp.9980–9984.

Dehestani, A. & Kazemi Tabar, S.K., 2007. A rapid efficient method for DNA isolation from plants with high levels of secondary metabolites. *Asian Journal of Plant Sciences*, 6(6), pp.977–981.

Dellaporta, S.L., Wood, J. & Hicks, J.B., 1983. A plant DNA minipreparation: Version II. *Plant Molecular Biology Reporter*, 1(4), pp.19–21.

Delong, J.M. & Steffen, K.L., 1997. Photosynthetic function, lipid peroxidation, and alpha-tocopherol content in spinach leaves during exposure to UV-B radiation. *Canadian Journal of Plant Science*, 77, pp.453–459.

Demidchik, V., 2015. Mechanisms of oxidative stress in plants: From classical chemistry to cell biology. *Environmental and Experimental Botany*, 109, pp.212–228.

Deng, J. et al., 2016. Cullin 4A (CUL4A), a direct target of miR-9 and miR-137, promotes

gastric cancer proliferation and invasion by regulating the Hippo signalling pathway. *Oncotarget*, 7(9), pp.10037–50.

Desai, P. et al., 2015. Comparative assessment of genetic diversity among Indian bamboo genotypes using RAPD and ISSR markers. *Molecular Biology Reports*, 42(8), pp.1265–1273.

Desgagné-Penix, I. et al., 2010. Integration of deep transcriptome and proteome analyses reveals the components of alkaloid metabolism in *Opium poppy* cell cultures. *BMC plant biology*, 10(252), pp.1-17.

Deshmukh, R. et al., 2014. Integrating omic approaches for abiotic stress tolerance in soybean. *Frontiers in Plant Science*, 5(June), pp.1-12.

Desmarchelier, C. et al., 1997. Total reactive antioxidant potential (TRAP) and total antioxidant reactivity (TAR) of medicinal plants used in Southwest Amazonia (Bolivia and Peru). *International Journal of Pharmacognosy*, 35(4), pp.288–296.

Devos, K.M. et al., 1993. Chromosomal rearrangements in the rye genome relative to that of wheat. *Theoretical and Applied Genetics*, 85(6–7), pp.673–680.

Dhindsa, R.S., Plumb-dhindsa, P. & Thorpe, T.A., 1981. Leaf senescence: Correlated with increased levels of membrane permeability and lipid peroxidation, and decreased levels of superoxide dismutase and catalase leaf senescence: Membrane decreased catalase correlated with increased levels of and permeability. *Journal of Experimental Botany*, 32(1), pp.93–101.

Diederichsen, A. & Hammer, K., 2003. The infraspecific taxa of coriander (*Coriandrum sativum* L.). *Genetic Resources and Crop Evolution*, 50, pp.33–63.

Dirisala, V.R. et al., 2017. Recombinant pharmaceutical protein production in plants: unravelling the therapeutic potential of molecular pharming. *Acta Physiologiae Plantarum*, 39(18), pp.1-9.

Dit Frey, N.F. et al., 2010. The RNA binding protein Tudor-SN is essential for stress tolerance and stabilises levels of stress-responsive mRNAs encoding secreted proteins in *Arabidopsis*. *The Plant Cell*, 22(5), p.1575-1591.

Dixon, R.A., Xie, D.Y. & Sharma, S.B., 2005. Proanthocyanidins - A final frontier in flavonoid research? *New Phytologist*, 165(1), pp.9–28.

Domínguez, F., González, M.C. & Cejudo, F.J., 2002. A germination-related gene encoding a serine carboxypeptidase is expressed during the differentiation of the vascular tissue in wheat grains and seedlings. *Planta*, 215(5), pp.727–734.

Domon, B. & Aebersold, R., 2006. Mass spectrometry and protein analysis. *Science (New York)*, 312(5771), pp.212–7.

Donaldson, J.S. et al., 2003. Cycads in trade and sustainable use of cycad populations. In Donaldson, J. ed. *Cycads: status survey and conservation action plan*. Gland, Switzerland and Cambridge, UK: IUCN, pp.48-53.

Dong, X. et al., 2011. Effects of combined high-pressure homogenization and enzymatic treatment on extraction yield, hydrolysis and function properties of peanut proteins. *Innovative Food Science & Emerging Technologies*, 12(4), pp.478–483.

Doolan, I.A. & Wilkinson, M.G., 2009. Comparison of the effects of various attenuation methods on cell permeability and accessibility of intracellular enzymes in *Lactococcus lactis* strains. *International Dairy Journal*, 19(4), pp.215–221.

Doyle, J.L. & Doyle, J.J., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull*, 19, pp.11–15.

DRI Committee, 2005. Dietary Reference Intakes for water, potassium, sodium, chloride, and sulfate. I*nsitute of Medicina of the National Acedemies*. The National

Academies Press, Washington D.C.

Dubcovsky, J. et al., 2001. Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes. *Plant Physiology*, 125(3), pp.1342–53.

Dung, N.X. & Loi, D.T., 1991. Selection of traditional. *Journal of Ethnopharmacology*, 32, pp.57–70.

Dunwell, J.M. et al., 2008. Germin and germin-like proteins: Evolution, structure, and function. *Critical Reviews in Plant Sciences*, 27(5), pp.342–375.

e Silva, T.A.S. et al., 2011. Purification and some properties of an extracellular acid protease from *Aspergillus clavatus*. *World Journal of Microbiology and Biotechnology*, 27(11), pp.2491–2497.

Eddinger, T.J. & Meer, D.P., 2007. Myosin II isoforms in smooth muscle: heterogeneity and function. *AJP: Cell Physiology*, 293(2), pp.C493–C508.

Edreva, A. et al., 2008. Stress-protective role of secondary metabolites: Diversity of functions and mechanisms. *Gen. Appl. Plant Physiology*, 34(1–2), pp.67–78.

Einsenthal, R. & Danson, M.J., 1992. *Enzyme assays, a practical approach*, Oxford University Press.

Ekblom, R. & Galindo, J., 2011. Applications of next-generation sequencing in molecular ecology of non-model organisms. *Heredity*, 107(1), pp.1–15.

El-Kassaby, Y., 1992. Domestication and genetic diversity-should we be concerned? *The Forestry Chronicle*, 68(6), pp.687–700.

El-Kassaby, Y.A. & Ritland, K., 1996. Impact of selection and breeding on the genetic diversity in Douglas-fir. *Biodiversity and Conservation*, 5(6), pp.795–813.

El-Mahdy, M.A. et al., 2006. Cullin 4A-mediated Proteolysis of DDB2 Protein at DNA Damage Sites Regulates in Vivo Lesion Recognition by XPC. *Journal of Biological Chemistry*, 281(19), pp.13404–13411.

Ellegren, H. & Ellegren, N., 2016. Determinants of genetic diversity. *Nature Reviews Genetics*, 17(7), pp.422–433.

Ellis, M. et al., 2010. Arabinogalactan-proteins: Key regulators at the cell surface? *Plant Physiology*, 153(2), pp.403-419.

Ellstrand, N.C., Ornduff, R. & Clegg, J.M., 1990. Genetic structure of the Australian cycad, *Macrozamia mommunis* (Zamiaceae). *American Journal of Botany*, 77(5), pp.677–681.

Emdad, L. et al., 2013. AEG-1/MTDH/LYRIC: Signaling pathways, downstream genes, interacting proteins, and regulation of tumour angiogenesis. In D. Sarkar & P. B. B. T.-A. in C. R. Fisher, eds. *AEG-1/MTDH/LYRIC Implicated in Multiple Human Cancers*. Academic Press, pp. 75–111.

Evans, L.M. et al., 2014. Population genomics of *Populus trichocarpa* identifies signatures of selection and adaptive trait associations. *Nature Genetics*, 46, pp.1089-1096.

Facchini, P.J. et al., 2012. Synthetic biosystems for the production of high-value plant metabolites. *Trends in Biotechnology*, 30(3), pp.127–131.

Fan, P.A.Z.L.D., 2013. Involvement of antioxidant defence system in chill hardening-induced chilling tolerance in *Jatropha curcas* seedlings. *Acta Physiologiae Plantarum*, 35, pp.153–160.

Fariss, M.W. & Reed, D.J., 1987. High-performance liquid chromatography of thiols and disulfides: Dinitrophenol derivatives. *Methods in Enzymology*, 143, pp.101–109.

Farmer, E.E. & Mueller, M.J., 2013. ROS-mediated lipid peroxidation and RES-activated signalling. *Annual Review of Plant Biology*, 64(1), pp.429–450.

Faye, L. & Gomord, V., 2010. Success stories in molecular farming - a brief overview. *Plant Biotechnology Journal*, 8(5), pp.525–528.

Fazeli, F., Ghorbanli, M. & Niknam, V., 2007. Effect of drought on biomass, protein content, lipid peroxidation and antioxidant enzymes in two sesame cultivars. *Biologia Plantarum*, 51(1), pp.98–103.

Feder, A. et al., 2015. A Kelch domain-containing F-box coding gene negatively regulates flavonoid accumulation in *Cucumis melo* L. *Plant Physiology*, 176(4), pp.1-32.

Felner, K.M. et al., 1998. Repressed beta-1,3-galactosyltransferase in the Tn syndrome. *Biochimica et biophysica Acta*, 1406(1), pp.115–125.

Fernandez, M. et al., 2003. Development of a phylogenetic tree in garlic (*Allium sativum* L.) using targeted mtDNA-PCR and RAPD analysis. *Biology*, 37, pp.51–54.

Figueiredo, M.S.L. & Grelle, C.E. V, 2009. Predicting global abundance of a threatened species from its occurrence: Implications for conservation planning. *Diversity and Distributions*, 15(1), pp.117–121.

Fíla, J. et al., 2012. Revealing phosphoproteins playing a role in tobacco pollen activated *in vitro*. *Proteomics*, 12(21), pp.3229–3250.

Fischer, C. et al., 2013. An IQ domain mediates the interaction with calmodulin in a plant cyclic nucleotide-gated channel. *Plant and Cell Physiology*, 54(4), pp.573–584.

Floegel, A. et al., 2011. Comparison of ABTS/DPPH assays to measure antioxidant capacity in popular antioxidant-rich US foods. *Journal of Food Composition and Analysis*, 24(7), pp.1043–1048.

Forman, H.J., 2016. Glutathione - from antioxidant to post-translational modifier. *Archives of Biochemistry and Biophysics*, 595, pp.64–67.

Forman, H.J., Zhang, H. & Rinna, A., 2009. Glutathione: Overview of its protective roles, measurement, and biosynthesis. *Molecular Aspects of Medicine*, 30(1–2), pp.1–12.

Franciosini, A. et al., 2013. The *Arabidopsis* COP9 signalosime interacting F-box Kelch 1 protein forms an SCF ubiquitin ligase and regulates hypocotyl elongation. *Molecular Plant*, 6(5), pp.1616–1629.

Frankham, R., Ballou, J. & Briscoe, D., 2010. *Introduction to Conservation Genetics* 2nd edn., Cambridge: University Press, Cambridge, UK.

Franzini, L. et al., 2012. Food selection based on high total antioxidant capacity improves endothelial function in a low cardiovascular risk population. *Nutrition, Metabolism, and Cardiovascular Diseases*, 22(1), pp.50–7.

Freitas, C.D.T. et al., 2017. Identification and characterization of two germin-like proteins with oxalate oxidase activity from *Calotropis procera* latex. *International Journal of Biological Macromolecules*, 105, pp.1051–1061.

Freitas, C.D.T. et al., 2016. Proteomic analysis and purification of an unusual germin-like protein with proteolytic activity in the latex of Thevetia peruviana. *Planta*, 243(5), pp.1115–1128.

Friso, G. et al., 2010. Reconstruction of metabolic pathways, protein expression, and homeostasis machinery across maize bundle sheath and mesophyll chloroplasts: large-scale quantitative proteomics using the first maize genome assembly. *Plant Physiol*, 152(3), pp.1219–1250.

Fu, J. et al., 2013. Genetic characterisation and authentication of *Lonicera japonica* Thunb. by using improved RAPD analysis. *Molecular Biology Reports*, 40(10), pp.5993–5999.

Fu, J.E. et al., 2016. Distinctive morphological characteristics contribute to the identification of *Artemisia annua* L. germplasms with high yield and high

artemisinin content. *Journal of Applied Research on Medicinal and Aromatic Plants*, 3(2), pp.43–47.

Fu, J.Y. et al., 2018. Identification and functional analysis of germin-like protein gene family in tea plant (*Camellia sinensis*). *Scientia Horticulturae*, 234, pp.166–175.

Fu, L. et al., 2011. Antioxidant capacities and total phenolic contents of 62 fruits. *Food Chemistry*, 129(2), pp.345–350.

Fu, X. et al., 2018. Oncoprotein Tudor-SN is a key determinant providing survival advantage under DNA damaging stress. *Cell Death & Differentiation*.

Fu, Y. et al., 2016. Quantitative proteomics reveals the central changes of wheat in response to powdery mildew. *Journal of Proteomics*, 130, pp.108–119.

Fuentes, P. et al., 2016. A new synthetic biology approach allows transfer of an entire metabolic pathway from a medicinal plant to a biomass crop. *eLife*, 5(e13664), pp.1–26.

Fukao, Y., 2015. Discordance between protein and transcript levels detected by selected reaction monitoring. *Plant Signaling & Behavior*, 10(5), pp.1-7.

Fukayama, H., Abe, R. & Uchida, N., 2010. SDS-dependent proteases induced by ABA and its relation to Rubisco and Rubisco activase contents in rice leaves. *Plant Physiology and Biochemistry*, 48(10–11), pp.808–812.

Fukudome, A. & Fukuhara, T., 2017. Plant dicer-like proteins: double-stranded RNA-cleaving enzymes for small RNA biogenesis. *Journal of Plant Research*, 130(1), pp.33–44.

Galindo-González, L. et al., 2017. LTR-retrotransposons in plants: Engines of evolution. *Gene*, 626, pp.14–25.

Gangopadhyay, G. et al., 2007. Sex detection of *Carica papaya* and *Cycas circinalis* in pre-flowering stage by ISSR and RAPD. *Current Science*, 92(4), pp.524–526.

Gassmann, W., Hinsch, M.E. & Staskawicz, B.J., 1999. The *Arabidopsis* RPS4 bacterial-resistance gene is a member of the TIR-NBS-LRR family of disease-resistance genes. *Plant Journal*, 20(3), pp.265–277.

Gasteiger, E. et al., 2005. Protein identification and analysis tools on the ExPASy server BT - The proteomics protocols handbook. In J. M. Walker, ed. Totowa, NJ: Humana Press, pp. 571–607.

Gaur, S. & Wadhwa, N., 2008. Alkaline protease from senesced leaves of invasive weed *Lantana camara*. *African Journal of Biotechnology*, 7(24), pp.4602–4608.

Gehlot, H.S., Purohit, A. & Shekhawat, N.S., 2005. Metabolic changes and protein patterns associated with adaptation to salinity in *Sesamum indicum* cultivars. *Journal of Cell and Molecular Biology*, 4, pp.31–39.

George, V.C., Dellaire, G. & Rupasinghe, H.P.V., 2017. Plant flavonoids in cancer chemoprevention: role in genome stability. *Journal of Nutritional BiocheBistry*, 45, pp.1–14.

Georgieva, K. et al., 2017. Antioxidant defence during desiccation of the resurrection plant *Haberlea rhodopensis*. *Plant Physiology and Biochemistry*, 114, pp.51–59.

Ghosh, D. & Xu, J., 2014. Abiotic stress responses in plant roots: a proteomics perspective. *Frontiers in Plant Science*, 5(6), pp.1–13.

Giavalisco, P. et al., 2005. Proteome analysis of *Arabidopsis thaliana* by two-dimensional gel electrophoresis and matrix-assisted laser desorption/ionisation-time of flight mass spectrometry. *Proteomics*, 5(7), pp.1902–13.

Gietler, M., 2016. Changes in the reduction state of ascorbate and glutathione, protein oxidation and hydrolysis leading to the development of dehydration intolerance in

Triticum aestivum L . seedlings. *Plant Growth Regulation*, 79, pp.287–297.

Gill, S.S. et al., 2013. Glutathione and glutathione reductase: A boon in disguise for plant abiotic stress defence operations. *Plant Physiology and Biochemistry*, 70, pp.204–212.

Gill, S.S. & Tuteja, N., 2010. Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. *Plant Physiology and Biochemistry*, 48(12), pp.909–930.

Gimode, D. et al., 2016. Identification of SNP and SSR markers in finger millet using next-generation sequencing technologies. *PLoS ONE*, 11(7), pp.1–23.

Gladwin, M.T., Wang, X. & Hogg, N., 2006. Methodological vexation about thiol oxidation versus S-nitrosation—A commentary on "An ascorbate-dependent artefact that interferes with the interpretation of the biotin-switch assay." *Free Radical Biology and Medicine*, 41(4), pp.557–561.

Goetz, H. et al., 2004. Comparison of selected analytical techniques for protein sizing, quantitation and molecular weight determination. *Journal of Biochemical and Biophysical Methods*, 60(3), pp.281–93.

Goldberg, A.L., 2003. Protein degradation and protection against misfolded proteins. *Nature*, 426(6968), pp.895–899.

Gong, F., Hu, X. & Wang, W., 2015. Proteomic analysis of crop plants under abiotic stress conditions: where to focus our research? *Frontiers in plant science*, 6(June), p.418.

Görg, A., Weiss, W. & Dunn, M.J., 2004. Current two-dimensional electrophoresis technology for proteomics. *Proteomics*, 4(12), pp.3665–85.

Govindaraj, M., Vetriventhan, M. & Srinivasan, M., 2015. Importance of genetic diversity assessment in crop plants and its recent advances: An overview of its analytical perspectives. *Genetics Research International*, 2015, pp.1–14.

Graça, I. et al., 2016. Functional characterisation of a chitinase class III (CgCHI3) and a glutathione S-transferase (CgGST) involved in *Casuarina glauca*–Frankia symbiosis BT - Biological nitrogen fixation and beneficial plant-microbe interaction. In F. González-Andrés & E. James, eds. Cham: Springer International Publishing, pp. 97–107.

Grandbastien, M.A., 2015. LTR retrotransposons, handy hitchhikers of plant regulation and stress response. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1849(4), pp.403–416.

Granier, F. & de Vienne, D., 1986. Silver staining of proteins: standardised procedure for two-dimensional gels bound to polyester sheets. *Analytical Biochemistry*, 155(1), pp.45–50.

Grattapaglia, D., Sederoff, R. & Carolina, N., 1994. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. *Genetics*, 137, pp.1121–1137.

Grimaud, F. et al., 2013. Exploring chloroplastic changes related to chilling and freezing tolerance during cold acclimation of pea (*Pisum sativum* L.). *Journal of Proteomics*, 80, pp.145–159.

Grover, C.E., Armel, S. & Wendel, F.J., 2012. Targeted sequence capture as a powerful tool for evolutionary analysis1. *American Journal of Botany*, 99(2), pp.312–319.

Guevara-Olvera, L. et al., 2012. Expression of a germin-like protein gene (CchGLP) from a geminivirus-resistant pepper (*Capsicum chinense* Jacq.) enhances tolerance to geminivirus infection in transgenic tobacco. *Physiological and Molecular Plant*

Pathology, 78, pp.45–50.

Guillemaut, P. & Mardchal-drouard, L., 1992. Isolation of plant DNA: A fast, inexpensive, and reliable method. *Plant Molecular Biology Reporter*, 10(1), pp.60–65.

Gülçin, İ., 2012. Antioxidant activity of food constituents: an overview. *Archives of Toxicology*, 86(3), pp.345–91.

Guo, L. et al., 2013. Cullin-RING ubiquitin ligase family in plant abiotic stress pathways. *Journal of Integrative Plant Biology*, 55(1), pp.21–30.

Gurnani, C. et al., 2014. Conservation of medicinal plant *Bryophyllum pinnatum* (Lam.) Kurz by *in vitro* nodal culture. *International Journal of Pharmaceutical Sciences and Research*, 5(6), pp.2406–2411.

Gutierrez-Beltran, E. et al., 2016. Tudor staphylococcal nuclease: Biochemistry and functions. *Cell Death and Differentiation*, 23(11), pp.1739–1748.

Gutierrez-Beltran, E. et al., 2015. Tudor staphylococcal nuclease links formation of stress granules and processing bodies with mRNA catabolism in *Arabidopsis*. *The Plant Cell*, 27(3), pp.926–943.

Gutteridge, J.M., 1995. Lipid peroxidation and antioxidants as biomarkers of tissue damage. *Clinical Chemistry*, 41, pp.1819–28.

Gutteridge, J.M.C., 1975. The use of standards for malonyldialdehyde. *Analytical Biochemistry*, 69, pp.518–526.

Habermann, B. et al., 2004. The power and the limitations of cross-species protein identification by mass spectrometry-driven sequence similarity searches. *Molecular & Cellular Proteomics*, 3(3), pp.238–249.

Hadi, H. El, Vettor, R. & Rossato, M., 2018. Vitamin E as a treatment for nonalcoholic fatty liver disease: Reality or Myth? *Antioxidants*, 7(1), p.12.

Hall, D.E. et al., 2011. An integrated genomic, proteomic and biochemical analysis of (+)-3-carene biosynthesis in Sitka spruce (*Picea sitchensis*) genotypes that are resistant or susceptible to white pine weevil. *Plant Journal*, 65(6), pp.936–948.

Ham, B.K. et al., 2018. Insights into plant phosphate sensing and signalling. *Current Opinion in Biotechnology*, 49, pp.1–9.

Hamilton, A.C., 2004. Medicinal plants, conservation and livelihoods. *Biodiversity*, pp.1477–1517.

Hampp, R. & De Filippis, L.F., 1=980. Plastid protease activity and prolamellar body transformation during greening. *Plant Physiol*, 65(4), pp.663–668.

Hamrick, J.L. & Godt, M.J.W., 1990. Allozyme diversity in plant species. In *Plant population genetics, breeding, and genetic resources.* Sunderland, Massachusetts: Sinauer Associates Inc., pp. 43–63.

Hamrick, J.L. & Godt, M.J.W., 1996. Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions: Biological Sciences*, 351(1345), pp.1291–1298.

Hanania, U.R.I., Velcheva, M. & Sahar, N., 2004. An improved method for isolating high-quality DNA from *Vitis vinifera* Nuclei. *Plant Molecular Biology Reporter*, 22, pp.173–177.

Hanh, P. et al., 2014. A survey of traditional medicinal plants used by K'ho people in the buffer zone of Chu Yang Sin national park, Vietnam. *Journal of Vietnamese Environment*, 6(3), pp.276–280.

Harashima, H. & Schnittger, A., 2010. The integration of cell division, growth and differentiation. *Current Opinion in Plant Biology*, 13(1), pp.66–74.

Hare, P.D., Cress, W.A. & Van Staden, J., 2001. The effects of exogenous proline and proline analogues on *in vitro* shoot organogenesis in *Arabidopsis*. *Plant Growth Regulation*, 34(2), pp.203–207.

Harish et al., 2014. Conservation genetics of endangered medicinal plant *Commiphora wightii* in Indian thar desert. *Gene*, 535(2), pp.266–272.

Havaux, M. et al., 2000. Photodamage of the photosynthetic apparatus and its dependence on the leaf developmental stage in the npq1 *Arabidopsi*s mutant deficient in the xanthophyll cycle enzyme violaxanthin de-epoxidase. *Plant Physiol*, 124, pp.273–284.

Hayat, S. et al., 2012. Role of proline under changing environments: a review. *Plant Signaling & Behavior*, 7(11), pp.1456–66.

Hayden, M.J. & Sharp, P.J., 2001. Targeted development of informative microsatellite (SSR) markers. *Nucl. Acids Res.*, 29(8), pp.1-6.

Healey, A. et al., 2014. Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. *Plant Methods*, 10(1), p.21.

Hedrick, P.W., 2005. A standardised genetic differentiation measure. *Evolution,* 59(8), pp.1633–1638.

Heider, B. et al., 2007. RAPD variation among North Vietnamese *Flemingia macrophylla* (Willd.) Kuntze ex Merr. accessions. In: Hawksworth D.L. & Bull A.T. eds, *Plant Conservation and Biodiversity*, Springer, Dordrecht, pp.43-57.

Hendricks, J., 1987. The Gondwanan *Cycas.* Encephalartos. *J. Cycad Soc. South. Africa.*, 10, pp.24–25.

Hernández, I., Alegre, L. & Munné-Bosch, S., 2004. Drought-induced changes in flavonoids and other low molecular weight antioxidants in *Cistus clusii* grown under Mediterranean field conditions. *Tree Physiology*, 24, pp.1303–1311.

Heuer, B., 2003. Influence of exogenous application of proline and glycinebetaine on growth of salt-stressed tomato plants. *Plant Science*, 165(4), pp.693–699.

Hill, K.D., Nguyen, H.T. & Loc, P.K., 2004. The genus Cycas (Cycadaceae) in Vietnam. *The Botanical Review*, 70(2), pp.134–193.

Hindman, B.R. & Virginia, W., 2015. *Myosin II Isoforms in Breast Cancer: Role of Contractility in Matrix Organization, Cell Migration, and Endothelial Barrier Function*. West Virginia University.

Hoang, T.M.L., De Filippis, L.F. & Le, X.T., 2009. Salt tolerance and screening for genetic changes in rice mutants after gamma irradiation using RAPD and microsatellite (RAMP) markers. *The Open Horticulture Journal*, 2(1), pp.62–69.

Hoban, S.M. et al., 2013. Bringing genetic diversity to the forefront of conservation policy and management. *Conservation Genetics Resources*, 5(2), pp.593–598.

Hodges, D.M. et al., 1999. The thiobarbituric anthocyanin for estimating lipid peroxidation in plant tissues containing and other interfering. *Planta*, 207, pp.604–611.

Hoffman, L. et al., 2010. Physiological changes during cold acclimation of perennial ryegrass accessions differing in freeze tolerance. *Crop Science*, 50, pp.1037–1047.

Hou, D., 1955. A revision of the genus *Celastrus. Annals of the Missouri Botanical Garden*, 42(3), pp.215–302.

Hougaard, A.B. et al., 2013. ESR spin trapping for characterisation of radical formation in *Lactobacillus acidophilus* NCFM and *Listeria innocua*. *Journal of Microbiological Methods*, 94(3), pp.205–212.

Hu, J., Rampitsch, C. & Bykova, N. V., 2015. Advances in plant proteomics toward improvement of crop productivity and stress resistance. *Frontiers in Plant Science*, 6(April), pp.1–15.

Hu, X.Q. et al., 2014a. A new macrocyclic lactone and a new quinoflavan from *Celastrus hindsii. Phytochemistry Letters*, 7, pp.169–172.

Hu, X.Q. et al., 2014b. Three new diphenylpropanes from *Celastrus hindsii*. *Archives of Pharmacal Research*, 37(11), pp.1411–1415.

Hu, Y., 2016. Regulatory concern polymerase chain reaction carryover contamination (PCR) carryover contamination. In Samadikuchaksaraei, A., eds. *Polymerase Chain Reaction for Biomedical Applications*. IntechOpen, pp. 57–68.

Hua, Y. et al., 2016. iTRAQ-based quantitative proteomic analysis of cultivated *Pseudostellaria heterophylla* and its wild-type. *Journal of Proteomics*, 139, pp.13–25.

Hua, Z. & Vierstra, R.D., 2011. The Cullin-RING ubiquitin-protein ligases. *Annual Review of Plant Biology*, 62(1), pp.299–334.

Huang, D., Ou, B. & Prior, R.L., 2005. The chemistry behind antioxidant capacity assays. *Journal of Agricultural and Food Chemistry*, 53(6), pp.1841–56.

Huang, H.-C. et al., 2000. A novel agarofuran sesquiterpene, celahin D from *Celastrus hindsii* Benth. *Chem. Pharm. Bull.*, 48(7), pp.1079–1080.

Huff, D.R., Peakall, R. & Smouse, P.E., 1993. RAPD variation within and among natural populations of outcrossing buffalograss *Buchloë dactyloides* Nutt.. *Theoretical and Applied Genetics,* 86(8), pp.927–34.

Hung, H. V., De Filippis, L.F. & Buckney, R., 2011. Population structure and genetic diversity of the rare and endangered *Sinocalamus mucclure* and *Markhamia stipulata* in Ba Be National Park, Vietnam. *Asian Journal of Plant Sciences*, 10, pp.312–322.

Hung, T.N. & Chi, V.L., 2014. Country status report on medicinal and aromatic plants in Vietnam. In *Expert Consultation on Promotion of Medicinal and Aromatic Plants in the Asia-Pacific Region*. pp. 218–226.

Hurkman, W.J., Fornari, C.S. & Tanaka, C.K., 1989. A comparison of the effect of salt on polypeptides and translatable mRNAs in roots of a salt-tolerant and a salt-sensitive cultivar of barley. *Plant Physiology*, 90(4), pp.1444–56.

Hurkman, W.J. & Tanaka, C.K., 1986. Solubilization of plant membrane proteins for analysis by two-dimensional gel electrophoresis. *Plant Physiology*, 81(3), pp.802–6.

Hussien, H., Salem, H. & Mekki, B.E., 2015. Ascorbate-glutathione and α-tocopherol triad enhances antioxidant systems in cotton plants grown under drought stress. *International Journal of ChemTech Research*, 8(4), pp.1463–1472.

Ibrahim, H.M. & Jaafar, Z.H., 2013. Abscisic acid-induced changes in production of primary and secondary metabolites, photosynthetic capacity, antioxidant capability, antioxidant enzymes and lipoxygenase inhibitory activity of *Orthosiphon stamineus* Benth. *Molecules*, 18(7), pp.7957-7976.

Ibrahim, R.I.H., 2011. A modified CTAB protocol for DNA extraction from young flower petals of some medicinal plant species. *Geneconserve,* 10(40), pp.165-182.

Imin, N. et al., 2001. Characterisation of rice anther proteins expressed at the young microspore stage. *Proteomics*, 1(9), pp.1149–1161.

Indochina., B.I., 2004. *SourceBook of exiting [sic] and proposed protected areas in Vietnam.*, [Ha Noi]: BirdLife International in Indochina.

Ipek, M., 2003. Comparative analysis of genetic diversity in garlic (*Allium sativum* L.) using AFLP, RAPD, and isozyme markers and characterization of a cytoplasmic marker associated with the bolting phenotype. *PhD Thesis*, The University of Wisconsin - Madison.

Iqbal, N. et al., 2014. A new perspective of phytohormones in salinity tolerance: Regulation of proline metabolism. *Environmental and Experimental Botany*, 100, pp.34–42.

Irwandi, R.A. & Vacharaksa, A., 2016. The role of microRNA in periodontal tissue: A review of the literature. *Archives of Oral Biology*, 72, pp.66–74.

Isaacson, T. et al., 2006. Sample extraction techniques for enhanced proteomic analysis of plant tissues. *Nature Protocols*, 1(2), pp.769–74.

Ismail, N.Z. et al., 2016. Evaluation of genetic diversity of *Clinacanthus nutans* (Acanthaceae) using RAPD, ISSR and RAMP markers. *Physiology and Molecular Biology of Plants*, 22(4), pp.523–534.

Ivey, K.L. et al., 2017. Association of flavonoid-rich foods and flavonoids with risk of all-cause mortality. *British Journal of Nutrition*, 117(10), pp.1470–1477.

Jaakola, L. & Hohtola, A., 2010. Effect of latitude on flavonoid biosynthesis in plants. *Plant, Cell and Environment*, 33(8), pp.1239–1247.

Jacobs, D.I. et al., 2005. Proteome analysis of the medicinal plant C*atharanthus roseus. Planta*, 221(5), pp.690–704.

Jacobs, D.I., Heijden, R. Van Der & Verpoorte, R., 2000. Proteomics in plant biotechnology and secondary metabolism research. *Phytochem. Anal.*, 11(5), pp.277–287.

Jain, S.M. & Gupta, S.D., 2013. *Biotechnology of neglected and underutilised crops*. Springer Dordrecht, Heidelbeg, London, New York.

Jamwal, K., Bhattacharya, S. & Puri, S., 2018. Plant growth regulator mediated consequences of secondary metabolites in medicinal plants. *Journal of Applied Research on Medicinal and Aromatic Plants*, In Press, pp.1–13.

Jang, S.-M., Redon, C.E. & Aladjem, M.I., 2018. Chromatin-bound Cullin-Ring ligases: Regulatory roles in DNA replication and potential targeting for cancer therapy. *Frontiers in Molecular Biosciences*, 5(19), pp.1-19.

Jariwala, N. et al., 2017. Oncogenic role of SND1 in development and progression of hepatocellular carcinoma. *Cancer Research*, 77(12), pp.3306–3316.

Jensen, J., Matthieu, F. & Louis, B., 2016. The past, present and future of genomic scans for selection. *Molecular Ecology*, 25(1), pp.1–4.

Jia, H. et al., 2017. Increased affinity for RNA targets evolved early in animal and plant Dicer lineages through different structural mechanisms. *Molecular Biology and Evolution*, 34(12), pp.3047–3063.

Jia, L. et al., 2017. Dysregulation of CUL4A and CUL4B ubiquitin ligases in lung cancer. *Journal of Biological Chemistry*, 292(7), pp.2966–2978.

Jiang, J. & Xiong, Y.L., 2016. Natural antioxidants as food and feed additives to promote health benefits and quality of meat products: A review. *MESC*, 120, pp.107–117.

Jiang, T. et al., 2010. Genetic linkage maps of white birches (*Betula platyphylla* Suk. and *B. pendula* Roth) based on RAPD and AFLP markers. *Molecular Breeding*, 27(3), pp.347–356.

Jiang, X. et al., 2017. Comparison of phenolic compound accumulation profiles in eight evergreen woody core eudicots indicating the diverse ecological adaptability of *Camellia sinensis. Scientia Horticulturae*, 219, pp.200–206.

Jiang, Y. et al., 2018. Genetic diversity and structure analysis of the endangered plant species *Horsfieldia hainanensis* Merr. in China. *Biotechnology & Biotechnological Equipment*, 32(1), pp.95–101.

Jie, Y. et al., 2013. SND1 affects proliferation of hepatocellular carcinoma cell line SMMC-7721 by regulating IGFBP3 expression. *The Anatomical Record*, 296(10), pp.1568–1575.

Jiménez-Aguilar, D.M. & Grusak, M.A., 2017. Minerals, vitamin C, phenolics, flavonoids and antioxidant activity of *Amaranthus* leafy vegetables. *Journal of Food Composition and Analysis*, 58, pp.33–39.

Jorge, T.F. et al., 2009. Mass spectrometry-based plant metabolomics: Metabolite responses to abiotic stress. *Indian Journal of Experimental Biology*, 47(12), pp.987–992.

Jorrín-Novo, J. V. et al., 2015. Fourteen years of plant proteomics reflected in proteomics: Moving from model species and 2DE-based approaches to orphan species and gel-free platforms. *Proteomics*, 15(5–6), pp.1089–1112.

Jorrin-Novo, J. V, 2014. Plant proteomics methods and protocols. In J. V Jorrin-Novo et al., eds. *Plant Proteomics: Methods and Protocols*. Totowa, NJ: Humana Press, pp. 3–13.

Joshi, B.C. & Joshi, R.K., 2014. The role of medicinal plants in livelihood improvement in Uttarakhand. *International Journal of Herbal Medicine*, 1(6), pp.55–58.

Jost, L., 2009. D vs. GST: Response to Heller and Siegismund (2009) and Ryman and Leimar (2009). *Molecular Ecology*, 18(10), pp.2088–2091.

Jost, L., 2008. GST and its relatives do not measure differentiation. *Molecular Ecology*, 17(18), pp.4015–4026.

Jouhet, J., Maréchal, E. & Block, M.A., 2007. Glycerolipid transfer for the building of membranes in plant cells. *Progress in Lipid Research*, 46(1), pp.37–55.

Kabir, S. et al., 2014. Molecular characterisation of genetic diversity of mint gene pool through RAPD markers. *International Journal of Agriculture Innovation and Research*, 2(4), pp.576–580.

Kaczmarek, K. et al., 2002. Feature-based fuzzy matching of 2D gel electrophoresis images. *Journal of Chemical Information and Computer Sciences*, 42(6), pp.1431–42.

Kamiloglu, S. et al., 2016. A review on the effect of drying on antioxidant potential of fruits and vegetables. *Critical Reviews in Food Science and Nutrition*, 56(sup1), pp.110-129.

Kanazawa, K. et al., 2012. Short photoirradiation induces flavonoid synthesis and increases its production in postharvest vegetables. *Journal of Agricultural and Food Chemistry*, 60(17), pp.4359–4368.

Kang, J. et al., 2008. NMR-based metabolomics approach for the differentiation of ginseng (P*anax ginseng*) roots from different origins. *Archives of Pharmacal Research*, 31(3), pp.330–336.

Kanwischer, M. et al., 2005. Alterations in tocopherol cyclase activity in transgenic and mutant plants of A*rabidopsis* affect tocopherol content, tocopherol composition, and oxidative stress. *Plant Physiology*, 137(2), pp.713–723.

Karlsson, P.M. et al., 2015. The *Arabidopsis* thylakoid transporter PHT4;1 influences phosphate availability for ATP synthesis and plant growth. *Plant Journal*, 84(1), pp.99–110.

Kasprzewska, A., 2003. Plant chitinases - regulation and function. *Cellular and*

Molecular Biology Letters, 8(3), pp.809–824.

Kato, Y. et al., 2013. Possible compensatory role among chloroplast proteases under excess-light stress condition possible compensatory role among chloroplast proteases under excess-light stress condition. *Plant Signaling & Behavior*, 8(3), pp.e231981–e231983.

Katterman, F.R. & Shattuck, V.I., 1983. An effective method of DNA isolation from the mature leaves of *Gossypium* species that contain large amounts of phenolic terpenoids and tannins. *Preparative Biochemistry*, 13(4), pp.347–59.

Kaya, C. et al., 2007. Improved salt tolerance of melon (*Cucumis melo* L.) by the addition of proline and potassium nitrate. *Environmental and Experimental Botany*, 60(3), pp.397–403.

Kaya, M.D. et al., 2006. Seed treatments to overcome salt and drought stress during germination in sunflower (*Helianthus annuus* L.). *European Journal of Agronomy*, 24(4), pp.291–295.

Kazemi, M. & Hajizadeh, H.S., 2012. Assessment of genetic diversity of mints Iranian wild " Mentha aquatic " populations using RAPD marker. *Journal of Agriculture Technology*, 8(1), pp.327–336.

Kennedy, B.F. & De Filippis, L.F., 2004. Tissue degradation and enzymatic activity observed during protoplast isolation in two ornamental *Grevillea* species. *In Vitro Cellular & Developmental Biology-Plant*, 40(1), pp.119–125.

Keppel, G., Lee, S.W. & Hodgskiss, P.D., 2002. Evidence for long isolation among populations of a pacific cycad: genetic diversity and differentiation in *Cycas seemannii* A.Br. (Cycadaceae). *The Journal of Heredity*, 93(2), pp.133–139.

Kerzendorfer, C. et al., 2011. CUL4B-deficiency in humans: Understanding the clinical consequences of impaired Cullin 4-RING E3 ubiquitin ligase function. *Mechanisms of Ageing and Development*, 132(8–9), pp.366–373.

Khan, M.H. & Panda, S.K., 2008. Alterations in root lipid peroxidation and antioxidative responses in two rice cultivars under NaCl-salinity stress. *Acta Physiologiae Plantarum*, 30(1), pp.81–89.

Khan, S., Al-Qurainy, F. & Nadeem, M., 2012. Biotechnological approaches for conservation and improvement of rare and endangered plants of Saudi Arabia. *Saudi Journal of Biological Sciences*, 19(1), pp.1–11.

Khanna, P.K. et al., 2014. Correlation between morphological, chemical and RAPD markers for assessing genetic diversity in *Withania somnifera* (L .) Dunal. *J. Crop Sci. Biotech*, 17(1), pp.27–34.

Khatri, S., Saini, R. V & Chhillar, A.K., 2017. Molecular farming approach towards bioactive compounds BT - metabolic engineering for bioactive compounds: Strategies and processes. In V. C. Kalia & A. K. Saini, eds. Singapore: Springer Singapore, pp. 49–72.

Kidric, M., Kos, J. & Sabotic, J., 2014. Proteases and their endogenous inhibitors in the plant response to abiotic stress. *Botanica Serbica*, 38(1), pp.139–158.

Kim, E.-Y. et al., 2011. An improved method of protein isolation and proteome analysis with S*accharina japonica* (Laminariales) incubated under different pH conditions. *Journal of Applied Phycology*, 23(1), pp.123–130.

Kim, S.W. et al., 2016. An Integrated biochemical, proteomics, and metabolomics approach for supporting medicinal value of *Panax ginseng* fruits. *Frontiers in Plant Science*, 7(July), pp.1–14.

Kingston-Smith, a H. & Foyer, C.H., 2000. Bundle sheath proteins are more sensitive to

oxidative damage than those of the mesophyll in maize leaves exposed to paraquat or low temperatures. *Journal of Experimental Botany*, 51(342), pp.123–130.

Kiss, A.K. & Naruszewicz, M., 2012. Polyphenolic compounds characterization and reactive nitrogen species scavenging capacity of *Oenothera paradoxa* defatted seed extracts. *Food Chemistry*, 131(2), pp.485–492.

Kivcak, B. & Akay, S., 2005. Quantitative determination of a -tocopherol in *Pistacia lentiscus*, *Pistacia lentiscus* var . chia, and *Pistacia terebinthus* by TLC-densitometry and colorimetry. *Fitoterapia*, 76, pp.62–66.

Kleinwächter, M. et al., 2015. Moderate drought and signal transducer induced biosynthesis of relevant secondary metabolites in thyme (*Thymus vulgaris*), greater celandine (*Chelidonium majus*) and parsley (*Petroselinum crispum)*. *Industrial Crops and Products*, 64(1), pp.158–166.

Kleinwächter, M. & Selmar, D., 2014. Influencing the product quality by applying drought stress during the cultivation of medicinal plants. *Physiological Mechanisms and Adaptation Strategies in Plants Under Changing Environment: Volume 1*, 42, pp.57–73.

Klose, J., 1975. Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik*, 26(3), pp.231–243.

Kluge, I. et al., 1997. S-Nitrosoglutathione in rat cerebellum: Identification and quantification by liquid chromatography-mass spectrometry. *Journal of Neurochemistry*, 69(6), pp.2599–2607.

Kobata, A. & Amano, J., 2005. Altered glycosylation of proteins produced by malignant cells, and application for the diagnosis and immunotherapy of tumours. *Immunology and Cell Biology*, 83(4), pp.429–439.

Kocsy, Â. et al., 2000. Genetic study of glutathione accumulation during cold hardening in wheat. *Planta*, 210, pp.295–301.

Kohen, R. & Nyska, A., 2002. Toxicologic pathology. *Toxicologic Pathology*, 30(6), pp.620–650.

Komiya, R., 2017. Biogenesis of diverse plant phasiRNAs involves a miRNA-trigger and Dicer-processing. *Journal of Plant Research*, 130(1), pp.17–23.

Konishi, H., Ishiguro, K. & Komatsu, S., 2001. A proteomics approach towards understanding blast fungus infection of rice grown under different levels of nitrogen fertilisation. *Proteomics*, 1(9), pp.1162–1171.

Von Konrat, M. et al., 2014. Early land plants today (ELPT): How many liverwort species are there? *Phytotaxa*, 9(1), p.22.

Koyama, K. et al., 2012. Light quality affects flavonoid biosynthesis in young berries of Cabernet Sauvignon grape. *Phytochemistry*, 78, pp.54–64.

Kramer, A.T. & Havens, K., 2009. Plant conservation genetics in a changing world. *Trends in Plant Science*, 14(11), pp.599–607.

Kranner, I. & Birtic, S., 2005. A modulating role for antioxidants in desiccation tolerance 1. *Integr. Comp. Biol*, 45, pp.734–740.

Krimmel, B. et al., 2010. OH-radical induced degradation of hydroxybenzoic and hydroxycinnamic acids and formation of aromatic products - A gamma radiolysis study. *Radiation Physics and Chemistry*, 79(12), pp.1247–1254.

Krishnaiah, D., Sarbatly, R. & Nithyanandam, R., 2011. A review of the antioxidant potential of medicinal plant species. *Food and Bioproducts Processing*, 89(3), pp.217–233.

Krishnamurthy, V. et al., 2013. Conservation status and effects of harvest on an endemic multi-purpose cycad, *Cycas circinalis* L., Western Ghats, India. *Tropical Ecology*, 54(3), pp.309–320.

Krishnan, A. et al., 2015. Virus-induced gene silencing (VIGS) for elucidation of pathogen defence role of serine/threonine protein kinase in the non-model plant *Piper colubrinum* Link. *Plant Cell, Tissue and Organ Culture*, 122(2), pp.269–283.

Kroc, M. et al., 2014. New evidence of ancestral polyploidy in the Genistoid legume *Lupinus angustifolius* L. (narrow-leafed lupin). *Theoretical and Applied Genetics*, 127(5), pp.1237–1249.

Kukreja, S. et al., 2005. Plant water status, H_2O_2 scavenging enzymes, ethylene evolution and membrane integrity of. *Biologia Plantarum*, 49(2), pp.305–308.

Kumar, A. et al., 2014. Efficiency of ISSR and RAPD markers in genetic divergence analysis and conservation management of *Justicia adhatoda* L., a medicinal plant. *Plant Systematics and Evolution*, 300(6), pp.1409–1420.

Kumar, D. et al., 2017. Cross-kingdom regulation of putative miRNAs derived from happy tree in cancer pathway: A systems biology approach. *International Journal of Molecular Sciences*, 18(6), pp.1–21.

Kumar, D., Yusuf, M.A. & Singh, P., 2013. Modulation of antioxidant machinery in α-tocopherol-enriched transgenic *Brassica juncea* plants tolerant to abiotic stress conditions. *Protoplasma*, 250, pp.1079–1089.

Kumar, H. et al., 2016. RAPD and ISSR marker-based comparative evaluation of genetic diversity among Indian germplasms of *Euryale ferox*: an aquatic food plant. *Applied Biochemistry and Biotechnology*, 180(7), pp.1345–1360.

Kumar, J. et al., 2009. Genetic diversity analysis in *Cymbopogon* species using DNA markers. *OMICS: A Journal of Integrative Biology*, 2(1), pp.20–29.

Kumar, M. et al., 2017. Proteome analysis reveals extensive light stress-response reprogramming in the seagrass *Zostera muelleri* (Alismatales, Zosteraceae) metabolism. *Frontiers in Plant Science*, 7(2023), pp.1–19.

Kumar, S., 1999. Rapid isolation of DNA from dry and fresh samples of plants producing large amounts of secondary metabolites and essential oils. *Plant Molecular Biology Reporter*, 17, pp.1–7.

Kuo, L.M.Y. & Kuo, Y.H., 1997. Antitumour and anti-AIDS triterpenes from *Celastrus hindsii. Phytochemistry*, 44(7), pp.1275–1281.

Kurian, J.C., 2012. Ethno-medicinal plants of India, Thailand and Vietnam. *J Biodiversity*, 3(1), pp.61–75.

Kurth, E.G. et al., 2017. Myosin-driven transport network in plants. *Proceedings of the National Academy of Sciences*, 114(8), pp.E1385–E1394.

Kuruma, H. et al., 2009. Staphylococcal nuclease domain-containing protein 1 as a potential tissue marker for prostate cancer. *The American Journal of Pathology*, 174(6), pp.2044–2050.

Kuta, D.D. et al., 2005. Optimization of protocols for DNA extraction and RAPD analysis in West African fonio (*Digitaria exilis* and *Digitaria iburua*) germplasm characterisation. *African Journal of Biotechnology*, 4, pp.1368–1371.

Lambelet, P., Saucy, F. & Loliger, J., 1994. Radical exchange reactions between vitamin E, vitamin C and phospholipids in autoxidising polyunsaturated lipids. *Free Radical Res*, 20(1), pp.1–10.

Larsen, H.O. & Olsen, C.S., 2007. Unsustainable collection and unfair trade? Uncovering and assessing assumptions regarding Central Himalayan medicinal plant

conservation. *Biodiversity and Conservation*, 16(6), pp.1679–1697.

Lascano, H.R. et al., 2001. Antioxidant system response of different wheat cultivars under drought: field and *in vitro* studies. *Aust J. Plant Physiol*, 28, pp.1095–1102.

Laucou, V. et al., 1998. Genetic mapping in pea. 1. RAPD-based genetic linkage map of Pisum sativum. *Theoretical and Applied Genetics*, 97(5–6), pp.905–915.

Laxalt, A.M. & Munnik, T., 2002. Phospholipid signalling in plant defence. *Current Opinion in Plant Biology*, 5(4), pp.332–338.

Lehesranta, S.J. et al., 2006. Proteomic analysis of the potato tuber life cycle. *Proteomics*, 6(22), pp.6042–6052.

Leipner, J., Fracheboud, Y. & Stamp, P., 1999. Effect of growing season on the photosynthetic apparatus and leaf antioxidative defences in two maize genotypes of different chilling tolerance. *Environmental and Experimental Botany*, 42(42), pp.129–139.

Leng, L. & Zhang, D.E.X., 2011. Measuring population differentiation using GST or D? A simulation study with microsatellite DNA markers under a finite island model and nonequilibrium conditions. *Molecular Ecology*, 20(12), pp.2494–509.

Leśniewicz, K. et al., 2012. Plant plasma membrane-bound staphylococcal-like DNases as a novel class of eukaryotic nucleases. *BMC Plant Biology*, 12(195), pp.1-18.

Lewontin, R.C., 1972. Testing the Theory of Natural Selection. *Nature*, 236, pp.181-182.

Li, C.L. et al., 2018. Tudor staphylococcal nuclease is a structure-specific ribonuclease that degrades RNA at unstructured regions during microRNA decay. *RNA*, 24, pp.739-748.

Li, E. et al., 2015. Comparative proteomics analyses of two races of *Fusarium oxysporum f. sp. conglutinans* that differ in pathogenicity. *Scientific Reports*, 5, pp.1-21

Li, H.M. & Teng, Y.S., 2013. Transit peptide design and plastid import regulation. *Trends in Plant Science*, 18(7), pp.360–366.

Li, J.J. et al., 2012. Chemical constituents from the stems of *Celastrus orbiculatus*. *Chinese Journal of Natural Medicines*, 10(4), pp.279–283.

Li, M. et al., 2011. Mutations in the F-box gene improve the panicle architecture and enhance the grain yield in rice. *Plant Biotechnology Journal*, 9(9), pp.1002–1013.

Li, X. et al., 2014. Comparison of three common DNA concentration measurement methods. *Analytical Biochemistry*, 451, pp.18–24.

Li, X. et al., 2015. Sustainable utilisation of traditional Chinese medicine resources: Systematic evidence-based complementary and alternative medicine. *Evidence-Based Complementary and Alternative Medicine*, 2015, pp.1-10.

Li, Y. & Wang, X., 2017. The role of cullin4B in human cancers. *Experimental Hematology and Oncology*, 6(1), pp.2–9.

Li, Z. et al., 2016. Serine carboxypeptidase 46 regulates grain filling and seed germination in rice (*Oryza sativa* L.). *PLoS ONE*, 11(7), pp.1–18.

Liang, X.Q. et al., 2016. A new *Celastrus* species from the middle miocene of Yunnan, China and its palaeoclimatic and palaeobiogeographic implications. *Review of Palaeobotany and Palynology*, 225, pp.43–52.

Lichtenthaler, H.K., 2007. Biosynthesis, accumulation and emission of carotenoids, α-tocopherol, plastoquinone, and isoprene in leaves under high photosynthetic irradiance. *Photosynth Res*, 92, pp.163–179.

Lichtenthaler, H.K., 1979. Occurrence and function of prenyl lipids in the photosynthetic membrane. In L. A. Appelqvist & C. Liljenberg, eds. *Advances in the Biochemistry*

and Physiology of Plant Lipids. Elsevier/North Holland Biomedical Pr, pp. 57–78.

Lin, F., Qu, Y. & Zhang, Q., 2014. Phospholipids molecules regulating cytoskeletal organization in plant abiotic stress tolerance. *Plant Signaling & Behavior*, 9(5), pp.e28337(1-5).

Lindemann, C. et al., 2017. Strategies in relative and absolute quantitative mass spectrometry-based proteomics. *Bio. Chem.*, 398, pp.687–699.

Liska, A.J. & Shevchenko, A., 2003. Expanding the organismal scope of proteomics: Cross-species protein identification by mass spectrometry and its implications. *Proteomics*, 3(1), pp.19–28.

Liu, B. et al., 2007. *Oryza sativa* Dicer-like4 reveals a key role for small interfering RNA silencing in plant development. *The Plant Cell Online*, 19(9), pp.2705–2718.

Liu, C.L. et al., 2018. Expression of serine peptidase inhibitor Kunitz type 1 in differentiated thyroid cancer. *Histochemistry and Cell Biology*, pp.1-10.

Liu, C. et al., 2016. A role for leucoanthocyanidin reductase in the extension of proanthocyanidins. *Nature Plants*, 2, p.16182.

Liu, C., Yu, H. & Chen, S., 2011. Framework for sustainable use of medicinal plants in China. *Plant Diversity and Resources*, 33(151), pp.65–68.

Liu, L. et al., 2009. CUL4A abrogation augments DNA damage response and protection against skin carcinogenesis. *Molecular Cell*, 34(4), pp.451–460.

Liu, L.X., Xu, S.M. & Woo, K.C., 2005. Solar UV-B radiation on growth, photosynthesis and the xanthophyll cycle in tropical *Acacias* and *Eucalyptus*. *Environmental and Experimental Botany*, 54(2), pp.121–130.

Liu, M.E. et al., 2013. A functional polymorphism of PON1 interferes with microRNA binding to increase the risk of ischemic stroke and carotid atherosclerosis. *Atherosclerosis*, 228(1), pp.161–167.

Liu, R.H. & Finley, J., 2005. Potential cell culture models for antioxidant research. *Journal of Agricultural and Food Chemistry*, 53(10), pp.4311–4.

Liu, T.Y. et al., 2016. Identification of plant vacuolar transporters mediating phosphate storage. *Nature Communications*, 7, pp.1–11.

Liu, T.Y. et al., 2014. MicroRNA-mediated surveillance of phosphate transporters on the move. *Trends in Plant Science*, 19(10), pp.647–655.

Liu, W. & Zhu, J., 1997. Proline accumulation and salt-stress-induced gene expression in a salt-hypersensitive mutant of *Arabidopsis*. *Plant Physiology*, 114, pp.591–596.

Liu, Y. et al., 2013. Proanthocyanidin synthesis in *Theobroma cacao*: Genes encoding anthocyanidin synthase, anthocyanidin reductase, and leucoanthocyanidin reductase. *BMC Plant Biology*, 13(202), pp.1-19.

Loan, P.T., 2015. *Breeding and building plantation model of medicinal plant Celastrus hindsii Benth in Tuyen Quang Province*, Hung Vuong University. In Vietnamese.

Loan, P.T., 2012. Cutting propagation of *Celastrus hindssi. Journal of Science and Technology Hung Vuong University*, 24, pp.40–43.

Locatelli, M. et al., 2010. Total antioxidant activity of hazelnut skin (*Nocciola piemonte* PGI): Impact of different roasting conditions. *Food Chemistry*, 119(4), pp.1647–1655.

Lodhi, M. a. et al., 1994. A simple and efficient method for DNA extraction from grapevine cultivars and *Vitis* species. *Plant Molecular Biology Reporter*, 12(1), pp.6–13.

Long, Y. et al., 2015. Genetic analysis of litchi (*Litchi chinensis* Sonn.) in southern China by improved random amplified polymorphic DNA (RAPD) and inter-simple

sequence repeat (ISSR). *Molecular Biology Reports*, 42(1), pp.159–166.

Lowe, a J. et al., 2000. Conservation genetics of bush mango from central/west Africa: Implications from random amplified polymorphic DNA analysis. *Molecular Ecology*, 9(7), pp.831–41.

Lu, J.-J. et al., 2012. Genetic linkage maps of *Dendrobium moniliforme* and *D. officinale* based on EST-SSR, SRAP, ISSR and RAPD markers. *Scientia Horticulturae*, 137(0), pp.1–10.

Lü, J.-M. et al., 2010. Chemical and molecular mechanisms of antioxidants: experimental approaches and model systems. *Journal of Cellular and Molecular Medicine*, 14(4), pp.840–860.

Lu, S.C., 2011. Regulation of glutathione synthesis. *Molecular Aspects of Medicine*, 30(1–2), pp.42–59.

Lubbe, A. & Verpoorte, R., 2011. Cultivation of medicinal and aromatic plants for speciality industrial materials. *Industrial Crops and Products*, 34(1), pp.785–801.

Lukatkin, A.S., Anjum, N.A. & Mordovia, O., 2014. Control of cucumber (*Cucumis sativus* L .) tolerance to chilling stress - evaluating the role of ascorbic acid and glutathione. *Frontiers in Environmental Science*, 2(December), pp.1–6.

Lusby, W.R., 1988. *Celastrus angulatus. Journal of Natural Products*, 5(3), pp.537–542.

Ly, T.N., Shimoyamada, M. & Yamauchi, R., 2006. Isolation and characterisation of rosmarinic acid oligomers in *Celastrus hindsii* Benth leaves and their antioxidative activity. *Journal of Agricultural and Food Chemistry*, 54(11), pp.3786–93.

Ma, D. et al., 2014. Plant physiology and biochemistry expression of flavonoid biosynthesis genes and accumulation of flavonoid in wheat leaves in response to drought stress. *Plant Physiology et Biochemistry*, 80, pp.60–66.

Ma, F. et al., 2015. MiR-361-5p inhibits colorectal and gastric cancer growth and metastasis by targeting staphylococcal nuclease domain containing-1. *Oncotarget*, 6(19), pp.17404–17416.

Ma, K. et al., 2017. RNA interference of Dicer-1 and Argonaute-1 increasing the sensitivity of *Aphis gossypii* Glover (Hemiptera: Aphididae) to plant allelochemical. *Pesticide Biochemistry and Physiology*, 138, pp.71–75.

Macedo, M.L.R. et al., 2016. Antimicrobial activity of ILTI, a Kunitz-type trypsin inhibitor from *Inga laurina* (SW.) Willd. *Current Microbiology*, 72(5), pp.538–544.

Maciel, F.M. et al., 2011. Identification and partial characterization of two cysteine proteases from castor bean leaves (*Ricinus communis* L.) activated by wounding and methyl jasmonate stress. *Acta Physiologiae Plantarum*, 33(5), pp.1867–1875.

Mackey, A.J., Haystead, T.A.J. & Pearson, W.R., 2002. Getting more from less. *Molecular & Cellular Proteomics*, 1(2), pp.139–147.

Madadkar, M., Colville, L. & Smirnoff, N., 2014. Plant physiology and biochemistry: The induction of menadione stress tolerance in the marine microalga, *Dunaliella viridis*, through cold pretreatment and modulation of the ascorbate and glutathione pools. *Plant Physiology et Biochemistry*, 84, pp.96–104.

Madan, S. et al., 1995. Proline and proline metabolising enzymes in in-vitro selected NaCl-tolerant *Brassica juncea* L. under salt stress. *Annals of Botany*, 76(1), pp.51–57.

Madison, S.L. & Nebenführ, A., 2013. Understanding myosin functions in plants: Are we there yet? *Current Opinion in Plant Biology*, 16(6), pp.710–717.

Magel, E., Siegfried, H. & De Filippis, L.F., 2002. Random amplification of polymorphic DNA and reverse transcription polymerase chain reaction of RNA in studies of

sapwood and heartwood. In Chaffey, N., eds. *Wood Formation in Trees.* Taylor & Francis, pp. 319–337.

Maguire, T.L. & Sedgley, M., 1997. Genetic diversity in *Banksia* and *Dryandra* (Proteaceae) with emphasis on *Banksia cuneata*, a rare and endangered species. *Heredity*, 79(4), pp.394-401.

Maizura, M., Aminah, A. & Wan Aida, W.M., 2010. Total phenolic content and antioxidant activity of kesum (*Polygonum minus*), ginger (*Zingiber officinale*) and turmeric (*Curcuma longa*) extract. *International Food Research Journal*, 17, pp.45–53.

Manach, C. et al., 2005. Bioavailability and bioefficacy of polyphenols in humans. I. Review of 97 bioavailability studies. *The American Journal of Clinical Nutrition*, 81(1), p.230S–242S.

Mandal, A. et al., 2018. Identification, characterisation, expression profiling, and virus-induced gene silencing of armadillo repeat-containing proteins in tomato suggest their involvement in tomato leaf curl New Delhi virus resistance. *Functional & Integrative Genomics*, 18(2), pp.101–111.

Manen, J.-F. et al., 2005. A fully automatable enzymatic method for DNA extraction from plant tissues. *BMC Plant Biology*, 5, p.23.

Manjili, F.A., Sedghi, M. & Pessarakli, M., 2012. Effects of phytohormones on proline content and antioxidant enzymes of various wheat cultivars under salinity stress. *Journal of Plant Nutrition*, 35(7), pp.1098–1111.

Mann, M. & Jensen, O.N., 2003. Proteomic analysis of post-translational modifications. *Nature Biotechnology*, 21(3), pp.255–61.

Manna, S., 2015. An overview of pentatricopeptide repeat proteins and their applications. *Biochimie*, 113, pp.93–99.

Mantel, N., 1967. The detection of disease clustering and a generalised regression approach. *Cancer Res*, 27, pp.209–220.

Marcus, A., 1981. The biochemistry of plants : a comprehensive treatise / Vol. 6, Proteins and nucleic acids. *FEBS Letters,* 136(1), pp.187-188.

Marouga, R., David, S. & Hawkins, E., 2005. The development of the DIGE system: 2D fluorescence difference gel analysis technology. *Analytical and Bioanalytical Chemistry*, 382(3), pp.669–678.

Martin, G.B., Bogdanove, A.J. & Sessa, G., 2003. Understanding the functions of plants disease resistance proteins. *Annual Review of Plant Biology*, 54(1), pp.23–61.

Martin, G.B., Williams, J.G. & Tanksley, S.D., 1991. Rapid identification of markers linked to a *Pseudomonas* resistance gene in tomato by using random primers and near-isogenic lines. In *Proceedings of the National Academy of Sciences of the United States of America*. pp. 2336–40.

Martínez-Esteso, M.J., Sellés-Marchart, S., et al., 2011. A DIGE-based quantitative proteomic analysis of grape berry flesh development and ripening reveals key events in sugar and organic acid metabolism. *Journal of Experimental Botany*, 62(8), pp.2521–2569.

Martínez-Esteso, M.J., Casado-Vela, J., et al., 2011. iTRAQ-based profiling of grape berry exocarp proteins during ripening using a parallel mass spectrometric method. *Molecular bioSystems*, 7(3), pp.749–765.

Martínez-Esteso, M.J. et al., 2013. iTRAQ-based protein profiling provides insights into the central metabolism changes driving grape berry development and ripening. *BMC Plant Biology*, 13(1), p.167.

Martínez-Esteso, M.J. et al., 2015. The role of proteomics in progressing insights into plant secondary metabolism. *Frontiers in Plant Science*, 6(July), p.504.

Martínez-maqueda, D. et al., 2013. Proteomics in foods: Principle and application. In F. Toldrá & L. M. L. Nollet, eds. Boston, MA: Springer US, pp. 21–50.

Masumbuko, L.I., Sinje, S. & Kullaya, A., 2014. Genetic diversity and structure of East African tall coconuts in Tanzania using RAPD markers. *Open Journal of Genetics*, 4, pp.175–181.

Mathew, K.M. et al., 2014. Optimization of genomic DNA extraction from fresh and dry leaves of large cardamom (*Amomum subulatum* Roxb.) for diversity analysis. *Indian Journal of Biotechnology*, 13(2), pp.221–224.

Matsunaga, W. et al., 2015. A small RNA mediated regulation of a stress-activated retrotransposon and the tissue-specific transposition during the reproductive period in *Arabidopsis*. *Frontiers in Plant Science*, 6(February), pp.1–12.

McDonnell, E.M., Coughlan, S.J. & Wyn-Jones, R.G., 1983. Differential effects of abscisic acid on glycine betaine and proline accumulation in three plant species. *Zeitschrift fur Pflanzenphysiologie*, 109, pp.207–213.

McHale, L. et al., 2006. Plant NBS-LRR proteins: adaptable guards. *Genome Biology*, 7(4), p.212.

Mehta, S.K. & Gaur, J.P., 1999. Heavy metal-induced proline accumulation and its role in ameliorating metal toxicity in *Chlorella vulgaris*. *New Phytologist*, 143(2), pp.253–259.

Meijer, H.J.G. & Munnik, T., 2003. Phospholipid-based signalling in plants. *Annual Review of Plant Biology*, 54, pp.265–306.

Meirmans, P.G. & Hedrick, P.W., 2011. Assessing population structure: F(ST) and related measures. *Molecular Ecology Resources*, 11(1), pp.5–18.

Mekki, B.E., Hussien, H. & Salem, H., 2015. Role of glutathione, ascorbic acid and α-tocopherol in alleviation of drought stress in cotton plants. *International Journal of ChemTech Research*, 8(4), pp.1573–1581.

Meng, H. et al., 2017. Dicer and Argonaute genes involved in RNA interference. *Applied and Environmental Microbiology*, 83(7), pp.1–15.

Merken, H.M. & Beecher, G.R., 2000. Measurement of food flavonoids by high-performance liquid chromatography: A review. *Journal of Agricultural and Food Chemistry*, 48(3), pp.577–99.

Meyers, B.C., Morgante, M. & Michelmore, R.W., 2002. TIR-X and TIR-NBS proteins: Two new families related to disease resistance TIR-NBS-LRR proteins encoded in *Arabidopsis* and other plant genomes. *Plant Journal*, 32(1), pp.77–92.

Michiels, A. et al., 2003. Extraction of high-quality genomic DNA from latex-containing plants. *Analytical Biochemistry*, 315(1), pp.85–89.

Mierziak, J., Kostyn, K. & Kulma, A., 2014. Flavonoids as important molecules of plant interactions with the environment. *Molecules*, 19(10), pp.16240–16265.

Miettinen, K. et al., 2014. The seco-iridoid pathway from *Catharanthus roseus. Nature Communications*, 5, p.3606.

Miller, A.C. et al., 2014. Exploration of a rare population of Chinese chestnut in North America: stand dynamics, health and genetic relationships. *AoB PLANTS*, 6, pp.plu065-plu065.

Miller, G. et al., 2009. Unravelling 1-pyrroline-5-carboxylate-proline cycle in plants by uncoupled expression of proline oxidation enzymes. *Journal of Biological Chemistry*, 284(39), pp.26482–26492.

Min, W.K. et al., 2008. Reverse random amplified microsatellite polymorphism reveals enhanced polymorphisms in the 3' end of simple sequence repeats in the pepper genome. *Molecules and Cells*, 26(3), pp.250–257.

Minh, T. & Tran, H., 2005. Genetic variation in cultivated coffee (*Coffea arabica* L.) accessions in northern New South. *PhD Thesis*, Southern Cross University.

Minh, V. Van et al., 2014. Medicinal plants used by the Hre community in the Ba to district of central Vietnam. *Journal of Medicinal Plants Studies*, 2(3), pp.64–71.

Miret, J.A. & Munn, S., 2015. Redox signalling and stress tolerance in plants: a focus on vitamin E. *Annals of the New York Academy of Sciences*, 1340, pp.29–38.

Mishra, P., Bhoomika, K. & Dubey, R.S., 2013. Differential responses of antioxidative defence system to prolonged salinity stress in salt-tolerant and salt-sensitive Indica rice (*Oryza sativa* L.) seedlings. *Protoplasma*, (250), pp.3–19.

Mittler, R., 2002. Oxidative stress, antioxidants and stress tolerance. *Trends in Plant Science*, 7(9), pp.405–410.

Miyazaki, Y. et al., 2011. Overexpression of Lov Kelch protein *2* enhances cell elongation and increases cell number and ploidy in the hypocotyl of *Arabidopsis thaliana*. *Plant Biotechnology*, 28(3), pp.267–272.

Mlotshwa, S. et al., 2015. A novel chemopreventive strategy based on therapeutic microRNAs produced in plants. *Cell Research*, 25(4), pp.521–524.

Mocchegiani, E. et al., 2014. Vitamin E – gene interactions in ageing and age-related inflammatory diseases: Implications for treatment. A systematic review. *Ageing Research Reviews*, 14, pp.81–101.

Moghtaderi, H., Sepehri, H. & Attari, F., 2017. Combination of arabinogalactan and curcumin induces apoptosis in breast cancer cells *in vitro* and inhibits tumour growth via overexpression of p53 level *in vivo*. *Biomedicine and Pharmacotherapy*, 88, pp.582–594.

Mohammadi, A., Mansoori, B. & Baradaran, B., 2017. Regulation of miRNAs by herbal medicine: An emerging field in cancer therapies. *Biomedicine and Pharmacotherapy*, 86, pp.262–270.

Molassiotis, A. et al., 2013. Proteomics in the fruit tree science arena: New insights into fruit defence, development, and ripening. *Proteomics*, 13(12–13), pp.1871–1884.

Moloney, J.N. & Cotter, T.G., 2017. ROS signalling in the biology of cancer. *Seminars in Cell and Developmental Biology*, pp.1–15.

Mondini, L., Noorani, A. & Pagnotta, M. a., 2009. Assessing plant genetic diversity by molecular tools. *Diversity*, 1(1), pp.19–35.

Moniruzzaman, M. et al., 2014. Identification of phenolic acids and flavonoids in monofloral honey from Bangladesh by high-performance liquid chromatography: determination of antioxidant capacity. *BioMed Research International*, 2014, p.737490.

Ministry of Natural Resources and Environment, 2015. *Vietnam National Biodiversity strategy to 2020, vision to 2030*.

Moore, B.D. et al., 2014. Explaining intraspecific diversity in plant secondary metabolites in an ecological context. *The New Phytologist*, 201(3), pp.733–750.

Moravčíková, J. et al., 2017. Chitinase activities in wheat and its relative species. *Agriculture*, 63(1), pp.14–22.

Morina, F., Milić, S. & Jovanović, S.V., 2015. An improved HPLC-DAD method for simultaneously measuring phenolics in the leaves of *Tilia platyphyllos* and *Ailanthus altissima*. *Botanica Serbica*, 39(2), pp.177–186.

Morris, K., Thomas, H. & Rogers, L.J., 1996. Endopeptidases during the development and senescence of *Lolium temulentum* leaves. *Phytochemistry*, 41(2), pp.377–384.

Mosner, E. et al., 2012. Floodplain willows in fragmented river landscapes: Understanding spatio-temporal genetic patterns as a basis for restoration plantings. *Biological Conservation*, 153, pp.211–218.

Mu, X.-Y. et al., 2012. *Celastrus obovatifolius* sp. nov. (Celastraceae) from China. *Nordic Journal of Botany*, 30(1), pp.53–57.

Mu, X.-Y., Zhao, L.-C. & Zhang, Z.-X., 2017. Molecular analysis of Chinese *Celastrus* and *Tripterygium* and implications in medicinal and pharmacological studies. *Plos One*, 12(1), pp.1-11

Mu, X.Y., Zhao, L.C. & Zhang, Z.X., 2012. Phylogeny of *Celastrus* L. (Celastraceae) inferred from two nuclear and three plastid markers. *Journal of Plant Research*, 125(5), pp.619–630.

Muench, D.G., Zhang, C. & Dahodwala, M., 2012. Control of cytoplasmic translation in plants. *Wiley Interdisciplinary Reviews: RNA*, 3(2), pp.178–194.

Mugford, S.T. & Milkowski, C., 2012. Serine carboxypeptidase-like acyltransferases from plants. *Methods in Enzymology*, 516, pp.279–297.

Mugford, S.T. & Osbourn, A., 2010. Evolution of serine carboxypeptidase-like acyltransferases in the monocots. *Plant Signaling & Behavior*, 5(2), pp.193–195.

Munne-Bosch, S., 2005. The role of α-tocopherol in plant stress tolerance. *Journal of Plant Physiology*, 162, pp.743–748.

Munne´-Bosch, S. & Alegre, L., 2002. Interplay between ascorbic acid and lipophilic antioxidant defences in chloroplasts of water-stressed *Arabidopsis* plants. *FEBS Lett*, 524, pp.145–148.

Munnik, T., 2001. Phosphatidic acid: An emerging plant lipid second messenger. *Trends in Plant Science*, 6(5), pp.227–233.

Müntz, K., 2007. Protein dynamics and proteolysis in plant vacuoles. *Journal of Experimental Botany*, 58(10), pp.2391–2407.

Muoi, L.D. et al., 2009. *Vietnam Plant Resouce: Plants containing bioactive compounds*, Hanoi: Science and Technology Press. In Vietnamese.

Murray, M.G. & Thompson, W.F., 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research*, 8(19), pp.4321–4325.

Myouga, F. et al., 2018. Stable accumulation of photosystem II requires ONE-HELIX PROTEIN1 (OHP1) of the light harvesting-like family. *Plant Physiology*, 176(3).

Tel-Zur et al., 1999. Modified CTAB procedure for DNA isolation from *Epiphytic cacti* of the genera *Hylocereus* and *Selenicereus* (Cactaceae). *Plant Molecular Biology Reporter*, 17, pp.249–254.

Na, S., Payne, S.H. & Bandeira, N., 2016. Multi-species identification of polymorphic peptide variants via propagation in spectral networks. *Molecular & Cellular Proteomics*, 15(11), pp.3501–3512.

Naeem, M., Zainal, Z. & Ismail, I., 2015. Plant kelch containing F-box proteins: structure, evolution and functions. *RSC Adv.*, 5(53), pp.42808–42814.

Nag, A., Ahuja, P.S. & Sharma, R.K., 2015. Genetic diversity of high-elevation populations of an endangered medicinal plant. *AoB PLANTS*, 7.

Nagano, H. et al., 2014. Distinct substrate specificities of *Arabidopsis* DCL3 and DCL4. *Nucleic Acids Research*, 42(3), pp.1845–1856.

Nagappan, A. et al., 2012. Comparative root protein profiles of Korean ginseng (*Panax*

ginseng) and Indian ginseng (*Withania somnifera*). *The American Journal of Chinese Medicine*, 40(01), pp.203–218.

Nakabayashi, R. et al., 2014. Enhancement of oxidative and drought tolerance in *Arabidopsis* by overaccumulation of antioxidant flavonoids. *The Plant Journal*, 77, pp.367–379.

Nakabayashi, R. & Saito, K., 2015. Integrated metabolomics for abiotic stress responses in plants. *Current Opinion in Plant Biology*, 24, pp.10–16.

Nakanishi, K. & Yoshikawa, N., 2016. Immunoglobulin A nephropathies in children (includes HSP). *Pediatric Nephrology*, 28(1), pp. 1-62.

Nam, M.H. et al., 2005. Proteomic analysis of Korean ginseng (*Panax ginseng* C.A. Meyer). *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 815(1–2), pp.147–155.

Nam, V.Q., Thang, B. Van & Tho, N.T., 2013. Propagation of *Celastrus hindsii* Benth by tissue culture. *Tap chi Khoa hoc va cCng nghe Lam nghiep*, 2, pp.11–16. In Vietnamese.

Narain, P. & Darwin, C., 2000. Genetic diversity – conservation and assessment. *Current*, 79(2), pp.170–175.

Naru, J. et al., 2017. *Identification of differentially expressed proteins in retinoblastoma tumours using mass spectrometry-based comparative proteomic approach*, Failure: Problem in retrieving ISSN for S1874391917300556.

Narusaka, M. et al., 2016. Leucine zipper motif in RRS1 is crucial for the regulation of *Arabidopsis* dual resistance protein complex RPS4/RRS1. *Scientific Reports*, 6, pp.2–11.

Narusaka, M., Iuchi, S. & Narusaka, Y., 2017. Analyses of natural variation indicate that the absence of RPS4/RRS1 and amino acid change in RPS4 cause loss of their functions and resistance to pathogens. *Plant Signaling and Behavior*, 12(3), pp.1–8.

Naugžemys, D. et al., 2014. Phylogenetic analysis of the polymorphic 4× species complex *Lonicera caerulea* (Caprifoliaceae) using RAPD markers and noncoding chloroplast DNA sequences. *Biologia*, 69(5), pp.585–593.

Nawrot, R. et al., 2016. Combination of transcriptomic and proteomic approaches helps to unravel the protein composition of *Chelidonium majus* L. milky sap. *Planta*, 244(5), pp.1055–1064.

Nawrot, R. et al., 2017. Proteomic comparison of *Chelidonium majus* L. latex in different phases of plant development. *Plant Physiology and Biochemistry*, 112, pp.312–325.

Nawrot, R., Kalinowski, A. & Gozdzicka-Jozefiak, A., 2007. Proteomic analysis of *Chelidonium majus* milky sap using two-dimensional gel electrophoresis and tandem mass spectrometry. *Phytochemistry*, 68(12), pp.1612–1622.

Naz, S. et al., 2014. Molecular characterisation and phylogenetic relationship of different citrus varieties of Pakistan. *The Journal of Animal and Plant Sciences*, 24(1), pp.315–320.

Ndeh, D. & Gilbert, H.J., 2018. Biochemistry of complex glycan depolymerisation by the human gut microbiota. *FEMS Microbiology Reviews*, 42(2), pp.146–164.

Neale, D.B. & Kremer, A., 2011. Forest tree genomics: growing resources and applications. *Nature reviews. Genetics*, 12(2), pp.111–22.

Nebenführ, A. & Dixit, R., 2018. Kinesins and myosins: Molecular motors that coordinate cellular functions in plants. *Annual Review of Plant Biology*, 69(1), p.042817-040024.

Nei, M., 1973. Analysis of gene diversity in subdivided populations. *Proceedings of the*

National Academy of Sciences, 70(12), pp.3321-3323.

Nei, M., 1972. Genetic distance between populations. *Am Nat*, 106, pp.283–292.

Newton, R.P. et al., 2004. Plant proteome analysis by mass spectrometry: principles, problems, pitfalls and recent developments. *Phytochemistry*, 65(11), pp.1449–1485.

Ngwene, B. et al., 2017. Intercropping induces changes in specific secondary metabolite concentration in Ethiopian kale (*Brassica carinata*) and African nightshade (*Solanum scabrum*) under controlled conditions. *Frontiers in Plant Science*, 8(September), pp.1–12.

Ning, R. et al., 2015. Natural β-dihydroagarofuran-type sesquiterpenoids as cognition-enhancing and neuroprotective agents from medicinal plants of the genus *Celastrus*. *Journal of Natural Products*, 78(9), pp.2175–2186.

Nishimura, M.T. et al., 2017. TIR-only protein RBA1 recognizes a pathogen effector to regulate cell death in *Arabidopsis*. *Proceedings of the National Academy of Sciences*, 114(10), pp.E2053–E2062.

Noctor, G., Lelarge-Trouverie, C. & Mhamdi, A., 2015. The metabolomics of oxidative stress. *Phytochemistry*, 112(1), pp.33–53.

Noormohammadi, Z. et al., 2012. Genetic variation among Iranian pomegranates (*Punica granatu*m L.) using RAPD, ISSR and SSR markers. *Australian Journal of Crop Science*, 6(2), pp.268–275.

Nosáľová, G. et al., 2011. Antitussive and immunomodulating activities of instant coffee arabinogalactan-protein. *International Journal of Biological Macromolecules*, 49(4), pp.493–497.

Nussaume, L., 2011. Phosphate import in plants: focus on the PHT1 transporters. *Frontiers in Plant Science*, 2(November), pp.1–12.

O'Farrell, P.H., 1975. High resolution of proteins electrophoresis. *The Journal of Biological Chemistry*, 250(10), pp.4007–4021.

Ochatt, S. & Jain, S.M., 2007. *Breeding of neglected and under-utilised crops, spices and herbs.*, Enfield: Science Publishers, Inc.

Oda, K., 2012. New families of carboxyl peptidases: Serine-carboxyl peptidases and glutamic peptidases. *Journal of Biochemistry*, 151(1), pp.13–25.

Ohmiya, A. et al., 1998. Cloning of genes encoding auxin-binding proteins (ABP19_20) from peach_ significant peptide sequence similarity with germin-like proteins. *Plant and Cell Physiology*, 39(5), pp.492–499.

Oldham, J.T. et al., 2010. Shotgun proteomic analysis of yeast-elicited California poppy (*Eschscholzia californica)* suspension cultures producing enhanced levels of benzophenanthridine alkaloids. *Journal of Proteome Research*, 9(9), pp.4337–4345.

Oleszek, W. et al., 2002. Flavonoids from *Pinus sylvestris* needles and their variation in trees of different origin grown for nearly a century at the same area. *Biochemical Systematics and Ecology*, 30(11), pp.1011–1022.

Olmos, E. et al., 2017. Arabinogalactan proteins are involved in salt-adaptation and vesicle trafficking in tobacco by-2 cell cultures. *Frontiers in Plant Science*, 8(June), pp.1–14.

Omalsaad et al., 2014. Genetic relationship between roselle (*Hibiscus sabdariffa* L.) and kenaf (*Hibiscus cannabinus* L.) accessions through optimization of PCR based RAPD method. *Emirates Journal of Food and Agriculture*, 26(3), pp.247–258.

On, T. Van et al., 2001. A survey of medicinal plants in BaVi National Park, Vietnam: methodology and implications for conservation and sustainable use. *Biological Conservation*, 97, pp.295–304.

Ong, S., Foster, L.J. & Mann, M., 2003. Mass spectrometric-based approaches in quantitative proteomics. *Methods*, 29, pp.124–130.

Onoyovwe, A. et al., 2013. Morphine biosynthesis in opium poppy involves two cell types: sieve elements and laticifers. *The Plant cell*, 25(10), pp.4110–22.

Osborne, R. et al., 2007. *Cycads of Vietnam*, Singapore: Osborne, Roy.

Osborne, R. et al., 2010. *Cycas fugax*. The IUCN Red List of Threatened Species 2010: e.T42039A10635348.

Osborne, R., 1995. The world cycad census and a proposed revision of the threatened species status for cycad taxa. *Biological Conservation*, 71, pp.1–12.

Østerberg, J.T. et al., 2017. Accelerating the domestication of new crops: Feasibility and approaches. *Trends in Plant Science*, 22(5), pp.373–384.

Otsuka, K. et al., 2018. Maintaining good miRNAs in the body keeps the doctor away?: Perspectives on the relationship between food-derived natural products and microRNAs in relation to exosomes/extracellular vesicles. *Molecular Nutrition and Food Research*, 62(1), pp.1–12.

Ou, B. et al., 2002. Analysis of antioxidant activities of common vegetables employing oxygen radical absorbance capacity (ORAC) and ferric reducing antioxidant power (FRAP) assays: a comparative study. *Journal of Agricultural and Food Chemistry*, 50(11), pp.3122–3128.

Ounaroon, A. et al., 2003. (R,S)-Reticuline 7-O-methyltransferase and (R,S)-norcoclaurine 6-O-methyltransferase of *Papaver somniferum* - cDNA cloning and characterization of methyl transfer enzymes of alkaloid biosynthesis in opium poppy. *Plant Journal*, 36(6), pp.808–819.

Ozcan, G. et al., 2015. Preclinical and clinical development of siRNA-based therapeutics. *Advanced Drug Delivery Reviews*, 87, pp.108–119.

Palma, J.M. et al., 2002. Plant proteases, protein degradation, and oxidative stress: role of peroxisomes. *Plant Physiology and Biochemistry*, 40, pp.521–530.

Pan, S. et al., 2013. New perspectives on how to discover drugs from herbal medicines: CAM's outstanding contribution to modern therapeutics. *Evidence-Based Complementary and Alternative Medicine*, 2013.

Pathak, M.R. & Abido, M.S., 2014. The role of biotechnology in the conservation of biodiversity. *Journal of Experimental Biology and Agricultural Sciences*, 2(4), pp.353–363.

Pawłowski, T.A., 2007. Proteomics of European beech (*Fagus sylvatica* L.) seed dormancy breaking: Influence of abscisic and gibberellic acids. *Proteomics*, 7(13), pp.2246–2257.

Paz, T.A. et al., 2017. Proteome profiling reveals insights into secondary metabolism in *Maytenus ilicifolia* (Celastraceae) cell cultures producing quinone methide triterpenes. *Plant Cell, Tissue and Organ Culture*, 130(2), pp.405–416.

Pedreschi, R. et al., 2013. Post-harvest proteomics and food security. *Proteomics*, 13(12–13), pp.1772–1783.

Pereira, A.M., Pereira, L.G. & Coimbra, S., 2015. Arabinogalactan proteins: rising attention from plant biologists. *Plant Reproduction*, 28(1), pp.1–15.

Pérez-Farrera, M.A. et al., 2006. Demography of the cycad *Ceratozamia mirandae* (Zamiaceae) under disturbed and undisturbed conditions in a biosphere reserve of Mexico. *Plant Ecology*, 187(1), pp.97–108.

Perluigi, M., Coccia, R. & Butterfield, D.A., 2012. 4-Hydroxy-2-nonenal, a reactive product of lipid peroxidation, and neurodegenerative diseases: a toxic combination

illuminated by redox proteomics studies. *Antioxidants & Redox Signaling*, 17(11), pp.1590–609.

Permingeat, H.R., Romagnoli, M. V & Ruben, H., 1998. A simple method for isolating high yield and quality DNA from cotton (*Gossypium hirsutum* L.) leaves. *Plant Molecular Biology Reporter*, 16, pp.1–6.

Peterson, B.L. & Cummings, B.S., 2006. A review of chromatographic methods for the assessment of phospholipids in biological samples. *Biomedical Chromatography*, 20(3), pp.227–243.

Peterson, D.G., Boehm, K.S. & Stack, S.M., 1997. Isolation of milligram quantities of nuclear DNA from tomato. *Plant Molecular Biology Reporter*, 15(2), pp.148–153.

Petrussa, E. et al., 2013. Plant flavonoids-biosynthesis, transport and involvement in stress responses. *International Journal of Molecular Sciences*, 14(7), pp.14950–14973.

Pfeiffer, J. et al., 2006. Biosynthesis of flavan 3-ols by leucoanthocyanidin 4-reductases and anthocyanidin reductases in leaves of grape (*Vitis vinifera* L.), apple (*Malus domestica* Borkh.) and other crops. *Plant Physiology and Biochemistry*, 44(5–6), pp.323–334.

Phuc, T.X., 2009. Why did the forest conservation policy fail in the Vietnamese uplands? Forest conflicts in Ba Vi National Park in the Northern region. *International Journal of Environmental Studies,* 66(1), pp. 59-68.

Pichereaux, C. et al., 2015. Comparative shotgun proteomic analysis of wild and domesticated *Opuntia spp*. species shows a metabolic adaptation through domestication. *Journal of Proteomics*, 143, pp.353–364.

Picotti, P. et al., 2009. Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell*, 138(4), pp.795–806.

Pierik, R. et al., 2006. The Janus face of ethylene: growth inhibition and stimulation. *Trends in Plant Science*, 11(4), pp.176–183.

Pimm, S.L. et al., 2014. The biodiversity of species and their rates of extinction, distribution, and protection. *Science (New York)*, 344(6187), p.1246752.

Pimm, S.L. & Joppa, L.N., 2015. How many plant species are there, Where are they, and at What rate are they going extinct? *Annals of the Missouri Botanical Garden*, 100(3), pp.170–176.

Pinho, S.S. & Reis, C.A., 2015. Glycosylation in cancer: Mechanisms and clinical implications. *Nature Reviews Cancer*, 15(9), pp.540–555.

Pirŕ, S. et al., 2016. MicroRNA from *Moringa oleifera*: Identification by high throughput sequencing and their potential contribution to plant medicinal value. *PLoS ONE*, 11(3), pp.1–25.

Pirttilä, A.M. et al., 2001. DNA isolation methods for medicinal and aromatic plants. *Plant Molecular Biology Reporter*, 19(3), p.273a-f.

Pokharel, D. et al., 2016. The role of CD44 and ERM proteins in expression and functionality of P-glycoprotein in breast cancer cells. *Molecules*, 21(3), pp.1-14.

Pollastri, S. & Tattini, M., 2011. Flavonols: Old compounds for old roles. *Annals of Botany*, 108(7), pp.1225–1233.

Polytechnic, V., 1993. Antioxidant response and winter hardiness in red spruce (*Picea rubens* Sarg.). *New Phytologist*, 123, pp.365–374.

Pomastowski, P. & Buszewski, B., 2014. Two-dimensional gel electrophoresis in the light of new developments. *TrAC Trends in Analytical Chemistry*, 53, pp.167–177.

Poprac, P. et al., 2017. Targeting free radicals in oxidative stress-related human diseases.

Trends in Pharmacological Sciences, 38(7), pp.592–607.

Porebski, S., Bailey, L.G. & Baum, B.R., 1997. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Molecular Biology Reporter*, 15(1), pp.8–15.

Porth, I. & El-Kassaby, Y., 2014. Assessment of the genetic diversity in forest tree populations using molecular markers. *Diversity*, 6(2), pp.283–295.

Potters, G. et al., 2007. Stress-induced morphogenic responses: growing out of trouble? *Trends in Plant Science*, 12(3), pp.98–105.

Pourcel, L. et al., 2007. Flavonoid oxidation in plants: from biochemical properties to physiological functions. *Trends in Plant Science*, 12(1), pp.29–36.

Pourmorad, F., Hosseinimehr, S.J. & Shahabimajd, N., 2006. Antioxidant activity, phenol and flavonoid contents of some selected Iranian medicinal plants. *African Journal of Biotechnology*, 5(11), pp.1142–1145.

Poustini, K., Siosemardeh, A. & Ranjbar, M., 2007. Proline accumulation as a response to salt stress in 30 wheat (*Triticum aestivum* L.) cultivars differing in salt tolerance. *Genetic Resources and Crop Evolution*, 54(5), pp.925–934.

Prasad, N.L.R.M.N. V, 2007. Genetic diversity analysis of *Celastrus paniculatus* Willd. - a nearly threatened, cognitive and intelligence enhancer – by RAPD Markers. *Functional Plant Science and Biotechnology*, 1(1), pp.195-199.

Prasad, S., Gupta, S.C. & Tyagi, A.K., 2017. Reactive oxygen species (ROS) and cancer: Role of antioxidative nutraceuticals. *Cancer Letters*, 387, pp.95–105.

Price, S.J. et al., 2016. Kunitz trypsin inhibitor in addition to Bowman-Birk inhibitor influence stability of lunasin against pepsin-pancreatin hydrolysis. *Food Research International*, 90, pp.205–215.

Procházková, D., Boušová, I. & Wilhelmová, N., 2011. Antioxidant and prooxidant properties of flavonoids. *Fitoterapia*, 82(4), pp.513–23.

Puchooa, D., 2004. A simple, rapid and efficient method for the extraction of genomic DNA from lychee (*Litchi chinensis* Sonn.). *African Journal of Biotechnology*, 3(4), pp.253–255.

Puga, M.I. et al., 2017. Novel signals in the regulation of Pi starvation responses in plants: facts and promises. *Current Opinion in Plant Biology*, 39(1), pp.40–49.

Pyngrope, S., Bhoomika, K. & Dubey, R.S., 2013. Reactive oxygen species, ascorbate-glutathione pool, and enzymes of their metabolism in drought-sensitive and tolerant indica rice (*Oryza sativa* L.) seedlings subjected to progressing levels of water deficit. *Protoplasma*, 250(2), pp.585–600.

Qian-Hua, J., Ji-wu, Z. & Yan-jun, G., 2011. Using optimized random amplified polymorphic DNA (RAPD) markers to identify the category status of *Citrus nobilis* Lour. Gonggan. *African Journal of Biotechnology*, 10(64), pp.13982–13990.

Qin, W. et al., 2005. Peripheral B lymphocyte β1,3-galactosyltransferase and chaperone expression in immunoglobulin A nephropathy. *Journal of Internal Medicine*, 258(5), pp.467–477.

Qu, Y. et al., 2008. Identification of a novel group of putative *Arabidopsis thaliana* β-(1,3)-galactosyltransferases. *Plant Molecular Biology*, 68(1–2), pp.43–59.

Quiroga, A.M. et al., 2012. Water stress and abscisic acid exogenous supply produce differential enhancements in the concentration of selected phenolic compounds in Cabernet Sauvignon. *Journal of Berry Research*, 2(1), pp.33–44.

Rabilloud, T. et al., 2010. Two-dimensional gel electrophoresis in proteomics: Past, present and future. *Journal of Proteomics*, 73(11), pp.2064–2077.

Radha, P., Urla, R. & Reddy, V.D., 2015. Genetic diversity of Indian *Cycas* species assessed using RAPD markers. *American Journal of Research Communication*, 3(12), pp.95–105.

Rahman, I., Kode, A. & Biswas, S.K., 2007. Assay for quantitative determination of glutathione and glutathione disulfide levels using enzymatic recycling method. *Nat. Protocols*, 1(6), pp.3159–3165.

Rahman, K., 2007. Studies on free radicals, antioxidants, and co-factors. *Clinical Interventions in Aging*, 2(2), pp.219–36.

Rai, A., Saito, K. & Yamazaki, M., 2017. Integrated omics analysis of specialized metabolism in medicinal plants. *Plant Journal*, 90(4), pp.764–787.

Rai, S.K. & Mukherjee, A.K., 2010. Statistical optimization of production, purification and industrial application of a laundry detergent and organic solvent-stable subtilisin-like serine protease (Alzwiprase) from *Bacillus subtilis* DM-04. *Biochemical Engineering Journal*, 48(2), pp.173–180.

Rai, V.P. et al., 2013. Genetic diversity in capsicum germplasm based on microsatellite and random amplified microsatellite polymorphism markers. *Physiology and Molecular Biology of Plants*, 19(4), pp.575–586.

Raina, R., Chand, R. & Sharma, Y.P., 2011. Conservation strategies of some important medicinal plants. *Int. J. Med. Arom. Plants*, 1(3), pp.342–347.

Rajasekaran, D. et al., 2016. Staphylococcal nuclease and Tudor domain containing 1 (SND1 protein) promotes hepatocarcinogenesis by inhibiting monoglyceride lipase (MGLL). *Journal of Biological Chemistry*, 291(20), pp.10736–10746.

Ramakrishna, A. & Ravishankar, G.A., 2011. Influence of abiotic stress signals on secondary metabolites in plants. *Plant Signaling and Behavior*, 6(11), pp.1720–1731.

Rao, N.K., 2004. Plant genetic resources: Advancing conservation and use through biotechnology. *African Journal of Biotechnology*, 3(2), pp.136–145.

Rao, V.R. & Hodgkin, T., 2002. Genetic diversity and conservation and utilisation of plant genetic resources. *Plant Cell, Tissue and Organ Culture*, 68, pp.1–19.

Rasouli, M. & Karimi, R., 2015. Application of random amplified microsatellite polymorphism (RAMP) in *Prunus* characterisation and mapping. *Journal of Nuts*, 6(1), pp.1–5.

Rathore, A.S. et al., 2015. Chitinases from bacteria to human: Properties, applications, and future perspectives. *Enzyme Research*, 2015, pp.1–8.

Rausch, C. & Bucher, M., 2002. Molecular mechanisms of phosphate transport in plants. *Planta*, 216(1), pp.23–37.

Raven, J.A. et al., 2018. Costs of acquiring phosphorus by vascular land plants: patterns and implications for plant coexistence. *New Phytologist*, 217(4), pp.1420–1427.

Rawlings, N.D. et al., 2014. *MEROPS*: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research*, 42(D1), pp.D503–D509.

Rawlings, N.D., Barrett, A.J. & Bateman, A., 2011. Asparagine peptide lyases: A seventh catalytic type of proteolytic enzymes. *Journal of Biological Chemistry*, 286(44), pp.38321–38328.

Ray, S. et al., 2016. Turnabout is fair play: Herbivory-induced plant chitinases excreted in fall armyworm frass suppress herbivore defences in maize. *Plant Physiology*, 171(1), pp.694–706.

Raza, S.H. et al., 2007. Glycinebetaine-induced modulation of antioxidant enzymes activities and ion accumulation in two wheat cultivars differing in salt tolerance.

Environmental and Experimental Botany, 60(3), pp.368–376.

Ražná, K. et al., 2016. MicroRNA (miRNA) in food resources and medicinal plant. *Potravinarstvo*, 10(1), pp.188–194.

Reinheckel, T. et al., 1998. Comparison of protein oxidation and aldehyde formation during oxidative stress in isolated mitochondria. *Free Radical Research*, 29(4), pp.297–305.

Rejeb, K. Ben et al., 2015. Hydrogen peroxide produced by NADPH oxidases increases proline accumulation during salt or mannitol stress in *Arabidopsis thaliana*. *New Phytologist*, 208, pp.1138–1148.

Ren, F. et al., 2003. A comparative study of genetic relationships among the AA-genome *Oryza* species using RAPD and SSR markers. *Theoretical and Applied Genetics,* 108(1), pp.113–20.

Rena, A.B., 1975. Proline pyrroline-5-carboxylate reductase pumpkin activity. *Phytochemistry*, 14, pp.657–661.

Rengarajan, K. et al., 2002. Technical brief quantifying DNA concentrations using fluorometry: A comparison of fluorophores. *Molecular Vision*, 8, pp.416–421.

Ribeiro, A. et al., 2011. Actinorhizal plant defence-related genes in response to symbiotic Frankia. *Functional Plant Biology*, 38(8–9), pp.639–644.

Richardson, M.L., Keathley, C.P. & Peterson, C.L., 2016. Breeding system of the critically endangered Lakela's mint and influence of plant height on pollinators and seed output. *Population Ecology*, 58(2), pp.277–284.

Rio, A.H. del, Bamberg, J.B. & Huaman, Z., 1997. Assessing changes in the genetic diversity of potato gene banks. *Theoretical and Applied Genetics*, 95, pp.191–198.

Ritland, K. et al., 2005. Isozyme variation and the conservation genetics of Garry oak. *Can. J. Bot.*, 83, pp.1478–1487.

Robert, T. et al., 2015. Molecular insights into the pathogenesis of IgA nephropathy. *Trends in Molecular Medicine*, 21(12), pp.762–775.

Rocheta, M. et al., 2012. Corky, a gypsy-like retrotransposon is differentially transcribed in *Quercus suber* tissues. *BMC Research Notes*, 5, pp.1–6.

Rodriguez-Estrada, M.T., Mcclements, D.J. & Decker, E.A., 2011. Antioxidant and prooxidant activity behavior of phospholipids in stripped soybean oil-in-water emulsions. *J Am oil Chem Soc*, 88, pp.1409–1416.

Rodriguez-Moreno, L. et al., 2018. Tools of the crook- infection strategies of fungal plant pathogens. *Plant Journal*, 93(4), pp.664–674.

Rodziewicz, P. et al., 2014. Influence of abiotic stresses on plant proteome and metabolome changes. *Acta Physiologiae Plantarum*, 36(1), pp.1–19.

Rogers, S.O. & Bendich, A.J., 1985. Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. *Plant Molecular Biology*, 5(2), pp.69–76.

Rogstad, S.H. et al., 2001. DNA Extraction from plants: The use of pectinase. *Plant Molecular Biology Reporter*, 19, pp.353–359.

Romero-Rodríguez, M.C. et al., 2014. Improving the quality of protein identification in non-model species. Characterization of *Quercus ilex* seed and *Pinus radiata* needle proteomes by using SEQUEST and custom databases. *Journal of Proteomics*, 105, pp.85–91.

Ronoh, R. et al., 2018. Development of next-generation sequencing (NGS)-based SSRs in African nightshades: Tools for analysing genetic diversity for conservation and breeding. *Scientia Horticulturae*, 235, pp.152–159.

Roodbarkelari, F. et al., 2010. CULLIN 4-RING FINGER-LIGASE plays a key role in the control of endoreplication cycles in *Arabidopsis* trichomes. *Proceedings of the National Academy of Sciences*, 107(34), pp.15275–15280.

Roy-Macauley, H. et al., 2001. Aspartic protease in leaves of common bean (*Phaseolus vulgaris* L.) and cowpea (*Vigna unguiculata* L.): enzymatic activity, gene expression and relation to drought susceptibility. *FEBS Letters*, 492, pp.242–246.

Roy, D. et al., 1993. Counteraction of exogenous L-proline with NaCl in salt-sensitive cultivar of rice. *Biologia Plantarum*, 35(1), pp.69–72.

Roy, U.K. et al., 2018. Purification of lectin and Kunitz trypsin inhibitor from soya seeds. *Journal of Chromatographic Science*, 56(5), pp.436-442.

Rozov, S.M., Permyakova, N. V & Deineko, E. V, 2018. Main strategies of plant expression system glycoengineering for producing humanised recombinant pharmaceutical proteins. *Biochemistry (Moscow)*, 83(3), pp. 215-232.

Russo, G.L.R. et al., 2017. Ins and outs of flavonoids in cancer prevention vs. cancer therapy: A lesson from quercetin in leukaemia. *Proceedings*, 1(10), p.977.

Ryan, J.M. & Nebenführ, A., 2018. Update on myosin motors: Molecular mechanisms and physiological functions. *Plant Physiology*, 176, pp.119-127.

Sadak, M.S. et al., 2014. Role of ascorbic acid and α tocopherol in alleviating salinity stress on flax plant (*Linum usitatissimum* L.). *Journal of Stress Physiology & Biochemistry*, 10(1), pp.93–111.

Saeed, N., Khan, M.R. & Shabbir, M., 2012. Antioxidant activity, total phenolic and total flavonoid contents of whole plant extracts *Torilis leptophylla* L. *BMC Complementary and Alternative Medicine*, 12(1), p.221.

Saha, T., Bindu Roy, C. & Nazeer, M.A., 2005. Microsatellite variability and its use in the characterization of cultivated clones of *Hevea brasiliensis*. *Plant Breeding*, 124(1), pp.86–92.

Sahasrabudhe, A. & Deodhar, M., 2010. Standardization of DNA extraction and optimization of RAPD-PCR conditions in *Garcinia indica*. *International Journal of Botany*, 6(3), pp.293–298.

Saini, R.K. et al., 2014. Effect of dehydration methods on retention of carotenoids, tocopherols, ascorbic acid and antioxidant activity in *Moringa oleifera* leaves and preparation of a RTE product. *J Food Sci Technol*, 51(9), pp.2176–2182.

Saini, R.K. & Keum, Y., 2016. Tocopherols and tocotrienols in plants and their products: A review on methods of extraction, chromatographic separation, and detection. *Food Research International*, 82, pp.59–70.

Saito, H. & Ishihara, K., 1997. Antioxidant activity and active sites of phospholipids as antioxidants. *Journal of the American Oil Chemist's Society*, 74(12), pp.1531–1536.

Sakamoto, A. et al., 2015. *In vitro* and *in vivo* evidence for oxalate oxidase activity of a germin-like protein from azalea. *Biochemical and Biophysical Research Communications*, 458(3), pp.536–542.

Sala-Cirtog, M., Marian, C. & Anghel, A., 2015. New insights of medicinal plant therapeutic activity-The miRNA transfer. *Biomedicine and Pharmacotherapy*, 74, pp.228–232.

Saleh, B., 2015. Phylogenetic assessment of *Arthrocnemum macrostachyum* (Chenopodiaceae) genotypes, using RAMP markers. *Polish Botanical Journal*, 60(2), pp.293–299.

Samuilov, S. et al., 2016. Lead uptake increases drought tolerance of wild type and transgenic poplar (P*opulus tremula* x *P . alba*) overexpressing GSH. *Environmental*

Pollution, 216, pp.773–785.

Sanchez-Rodriguez, E. et al., 2010. Genotypic differences in some physiological parameters symptomatic for oxidative stress under moderate drought in tomato plants. *Plant Science*, 178(1), pp.30–40.

Sánchez de la Hoz, M.P. et al., 1996. Simple sequence repeat primers used in polymerase chain reaction amplification to study genetic diversity in barley. *Genome/National Research Council Canada*, 39(1), pp.112–7.

Sancho-Knapik, D. et al., 2017. Changes of secondary metabolites in *Pinus sylvestris* L. needles under increasing soil water deficit. *Annals of Forest Science*, 74(1).

Sangin, P., Thongpan, A. & Lindstrom, A.J., 2006. Molecular identification of *Cycas* by restriction fragment length polymorphism (RFLP) and random amplified polymorphic DNA (RAPD). *Kasetsart J. (Nat. Sci.)*, 120, pp.107–120.

Sanman, L.E. & Bogyo, M., 2014. Activity-based profiling of proteases. *The Annual Rev. Biochem*, 83.

Santos, D. et al., 2003. Genetic variation within and among species of five sections of the genus *Arachis* L. (Leguminosae) using RAPDs. *Genetic Resources and Crop Evolution*, 50, pp.841–848.

Santos, P. et al., 2010. Characterization of four defence-related genes up-regulated in root nodules of *Casuarina glauca*. *Symbiosis*, 50(1–2), pp.27–35.

Sapio, M.R. & Fricker, L.D., 2014. Carboxypeptidases in disease: Insights from peptidomic studies. *Proteomics - Clinical Applications*, 8(5–6), pp.327–337.

Saravanan, R.S. & Rose, J.K.C., 2004. A critical evaluation of sample extraction techniques for enhanced proteomic analysis of recalcitrant plant tissues. *Proteomics*, 4(9), pp.2522–32.

Sarikas, A., Hartmann, T. & Pan, Z.Q., 2011. The cullin protein family. *Genome Biology*, 12(4), pp.1–12.

Sarwat, M. et al., 2006. A standardised protocol for genomic DNA isolation from *Terminalia arjuna* for genetic diversity analysis. *Electronic Journal of Biotechnology*, 9(1), pp.86–91.

Sawant, R. & Nagendran, S., 2014. Protease: an enzyme with multiple industrial applications. *World Journal of Pharmacy and Pharmaceutical Sciences*, 3(6), pp.568–579.

Schaal, B.A. et al., 1998. Phylogeographic studies in plants. *Molecular Ecology*, pp.465–474.

Schaller, A., 2004. A cut above the rest: The regulatory function of plant proteases. *Planta*, 220(2), pp.183–197.

Schilmiller, A.L. et al., 2010. Studies of a biochemical factory: tomato trichome deep expressed sequence tag sequencing and proteomics. *Plant Physiology*, 153(3), pp.1212–1223.

Schippmann, U. et al., 2005. Impact of cultivation and collection on the conservation of medicinal plants: global trends and issues. In *III WOCMAP Congress on medicinal and aromatic plants: conservation, cultivation and sustainable use of medicinal and aromatic plants*. Chiang Mai.

Schmitz-Linneweber, C. & Small, I., 2008. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends in Plant Science*, 13(12), pp.663–670.

Schoettle, A.W. et al., 2012. Geographic patterns of genetic variation and population structure in *Pinus aristata* , Rocky Mountain bristlecone pine. *Canadian Journal of Forest Research*, 42(1), pp.23–37.

Schopp-Guth, A. & Fremuth, W., 2001. Sustainable use of medicinal plants and nature conservation in Prespa National Park area, Albania. *Medicinal Plants Conservation*, 7(2), pp.5–7.

Schou, M.F. et al., 2017. Unexpected high genetic diversity in small populations suggests maintenance by associative overdominance. *Molecular Ecology*, 26(23), pp.6510–6523.

Schubert, O.T. et al., 2017. Quantitative proteomics: challenges and opportunities in basic and applied research. *Nat. Protocols*, 12(7), pp.1289–1294.

Scoparo, C.T. et al., 2016. Chemical characterisation of heteropolysaccharides from green and black teas (*Camellia sinensis*) and their anti-ulcer effect. *International Journal of Biological Macromolecules*, 86, pp.772–781.

Scotti-Campos, P. et al., 2014. Phospholipids profile in chloroplasts of *Coffea spp.* genotypes differing in cold acclimation ability. *Journal of Plant Physiology*, 171(3–4), pp.243–249.

Seifert, G.J. & Roberts, K., 2007. The biology of arabinogalactan proteins. *Annual Review of Plant Biology*, 58(1), pp.137–161.

Semagn, K., Bjørnstad, Å. & Ndjiondjop, M.N., 2006. An overview of molecular marker methods for plants. *African Journal of Biotechnology*, 5(25), pp.2540–2568.

Semenova, T. et al., 2011. Evolutionary patterns of proteinase activity in attine ant fungus gardens. *BMC Microbiology*, 11(1), p.15.

Senapati, S.K., Aparajita, S. & Rout, G.R., 2013. Micropropagation and assessment of genetic stability in *Celastrus paniculatus*: An endangered medicinal plant. *Biologica*, 68(4), pp.627–632.

Seppanen, C.M., Song, Q. & Csallany, A.S., 2010. The antioxidant functions of tocopherol and tocotrienol homologues in oils, fats, and food systems. *J Am Oil Chem Soc*, 87, pp.469–481.

Serpen, A., Gökmen, V. & Fogliano, V., 2012. Total antioxidant capacities of raw and cooked meats. *Meat Science*, 90(1), pp.60–5.

Shakhtshneider, T.P. et al., 2016. New composites of betulin esters with arabinogalactan as highly potent anti-cancer agents. *Natural Product Research*, 30(12), pp.1382–1387.

Shao, J. et al., 2017. Aggregation of SND1 in stress granules is associated with the microtubule cytoskeleton during heat shock stimulus. *Anatomical Record*, 300(12), pp.2192–2199.

Sharma, A.D.E. V, Gill, P.K. & Singh, P., 2002. DNA isolation from dry and fresh samples of polysaccharide-rich plants. *Plant Molecular Biology Reporter*, 20, pp.415a–415f.

Sharma, K.K., Lavanya, M. & Anjaiah, V., 2000. A method for isolation and purification of peanut genomic DNA suitable for analytical applications. *Plant Molecular Biology Reporter*, 18, p.393a–393h.

Sharma, M. & Pandey, G.K., 2016. Expansion and function of repeat domain proteins during stress and development in plants. *Frontiers in Plant Science*, 6(1218), pp.1–15.

Sharma, S., Villamor, J.G. & Verslues, P.E., 2011. Essential role of tissue-specific proline synthesis and catabolism in growth and redox balance at low water potential. *Plant Physiology*, 157(1), pp.292–304.

Sheikh, K., Ahmad, T. & Khan, M.A., 2002. Use, exploitation and prospects for conservation: People and plant biodiversity of Naltar Valley, northwestern

Karakorums, Pakistan. *Biodiversity and Conservation*, 11(4), pp.715–742.

Shelly C. Lu, M.D., 2014. Glutathione synthesis. *Biochim Biophys Acta*, 1830(5), pp.3143–3153.

Shen, J. et al., 2013. Maximizing root/rhizosphere efficiency to improve crop productivity and nutrient use efficiency in intensive agriculture of China. *Journal of Experimental Botany*, 64(5), pp.1181–1192.

Sheoran, I.S. et al., 2009. Compatibility of plant protein extraction methods with mass spectrometry for proteome analysis. *Plant Science*, 176(1), pp.99–104.

Shepherd, L.D. & McLay, T.G.B., 2011. Two micro-scale protocols for the isolation of DNA from polysaccharide-rich plant tissue. *Journal of Plant Research*, 124(2), pp.311–4.

Shepherd, M. et al., 2002. High-throughput DNA extraction from forest trees. *Plant Molecular Biology*, 20, p.425a–425j.

Shi, J. et al., 2017. Differential proteomic analysis by iTRAQ reveals the mechanism of *Pyropia haitanensis* responding to high-temperature stress. *Nature Publishing Group*, *Scientific Report*, 7(44733), pp.1–11.

Shi, L. et al., 2018. Proanthocyanidin synthesis in Chinese bayberry (*Myrica rubra* Sieb. et Zucc.) fruits. *Frontiers in Plant Science*, 9(February), pp.1–12.

Shiba, S.N.S. et al., 2016. Exposing the illegal trade in encephalartos species at the Faraday 'muthi' market in South Africa using DNA barcoding. *South African Journal of Botany*, 1(1), p.348.

Showalter, A.M. & Basu, D., 2016. Glycosylation of arabinogalactan-proteins essential for development in *Arabidopsis*. *Communicative and Integrative Biology*, 9(3), pp.1–6.

Shrestha, C.L. et al., 2008. Chitinase levels in rice cultivars correlate with resistance to the sheath blight pathogen *Rhizoctonia solani*. *European Journal of Plant Pathology*, 120(1), pp.69–77.

Shulaev, V. et al., 2008. Metabolomics for plant stress response. *Physiologia Plantarum*, 132(2), pp.199–208.

Silva, P.A. et al., 2016. Leaf gas exchange and multiple enzymatic and non-enzymatic antioxidant strategies related to drought tolerance in two oil palm hybrids. *Trees-Structure and Function*, 30(1), pp.203–214.

Simmons, M.P. et al., 2012. Molecular phylogenetics and evolution phylogeny of Celastraceae tribe Euonymeae inferred from morphological characters and nuclear and plastid genes. *Molecular Phylogenetics and Evolution*, 62(1), pp.9–20.

Simmons, M.P. et al., 2008. Molecular phylogenetics and evolution phylogeny of the Celastreae and the relationships of Cathaedulis inferred from morphological characters and nuclear and plastid genes. *Molecular Phylogenetics and Evolution*, 48, pp.745–757.

Simon, E.W., 1974. Phospholipids and plant membrane permeability. *New Phytologist*, 73(3), pp.377–420.

Simova-Stoilova, L. et al., 2010. Proteolytic activity and cysteine protease expression in wheat leaves under severe soil drought and recovery. *Plant Physiology and Biochemistry*, 48(2–3), pp.200–206.

Singh, M. & Ahuja, P.S., 1999. Isolation and PCR amplification of genomic DNA from market samples of dry tea. *Plant Molecular Biology Reporter*, 17, pp.171–178.

Singh, N. & Sharma, A., 2017. Turmeric (*Curcuma longa*): miRNAs and their regulating targets are involved in development and secondary metabolite pathways. *Comptes*

Rendus - Biologies, 340(11–12), pp.481–491.

Singh, N., Srivastava, S. & Sharma, A., 2016. Identification and analysis of miRNAs and their targets in ginger using bioinformatics approach. *Gene*, 575(2), pp.570–576.

Singh, V., Nath, B. & Vinay, T., 2016. Interaction of Mg with heavy metals (Cu, Cd) in *T. aestivum* with special reference to oxidative and proline metabolism. *Journal of Plant Research*, 129(3), pp.487–497.

Sinha, R. & Chattopadhyay, S., 2011. Changes in the leaf proteome profile of *Mentha arvensis* in response to *Alternaria alternate* infection. *Journal of Proteomics*, 74(3), pp.327–336.

Siol, M., Wright, S.I. & Barrett, S.H., 2010. The population genomics of plant adaptation. *New Phytologist*, 188(2), pp.313–332.

Sjuts, I., Soll, J. & Bölter, B., 2017. Import of soluble proteins into chloroplasts and potential regulatory mechanisms. *Frontiers in Plant Science*, 8(February), pp.1–15.

Skaar, I. et al., 2014. Purple anthocyanin colouration on lower (abaxial) leaf surface of *Hemigraphis colorata* (Acanthaceae). *Phytochemistry*, 105, pp.141–146.

Slama, I. et al., 2015. Water deficit stress applied only or combined with salinity affects physiological parameters and antioxidant capacity in *Sesuvium portulacastrum*. *Flora - Morphology, Distribution, Functional Ecology of Plants*, 213, pp.69–76.

Slover, H.T., Thompson, R.H. & Merola, G. V, 1983. Determination of tocopherols and sterols by capillary gas chromatography. *J Am Oil Chem Soc*, 60(8), pp.1524–1528.

Sobrino-Plata, J. et al., 2014. The role of glutathione in mercury tolerance resembles its function under cadmium stress in *Arabidopsis*. *Metallomics*, 6, pp.356–366.

Sommerville, K., 2001. *Analysis of genetic variation in Grevillea caleyi R.BR using random amplified microsatellite polymorphisms (RAMPs)*. University of Technology, Sydney.

Son, N.T. & Binh, N.D., 2015. Effect of land-use change on run off and sediment yield in Da River Basin of Hoa Binh province, Northwest Vietnam. *J. Mt. Sci.*, 12(2), pp.1051–1064.

Song, Y. et al., 2012. DMSO, an organic cleanup solvent for TCA/Acetone-precipitated proteins, improves 2-DE protein analysis of rice roots. *Plant Molecular Biology Reporter*, 30(5), pp.1204–1209.

Spiro, R.G., 2002. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology*, 12(4), p.43R–56R.

Srivastava, S.K. et al., 2011. MicroRNA-150 directly targets MUC4 and suppresses growth and malignant behaviour of pancreatic cancer cells. *Carcinogenesis*, 32(12), pp.1832–1839.

Stajner, D. et al., 2006. Exploring *Allium* species as a source of potential medicinal agents. *Phytotherapy Research*, 20(7), pp.581–584.

Stalikas, C.D., 2007. Extraction, separation, and detection methods for phenolic acids and flavonoids. *Journal of Separation Science*, 30(18), pp.3268–95.

Stanjner, D. et al., 2011. Comparative study on *Allium schoenoprasum* cultivated plant and *Allium schoenoprasum* tissue culture organs antioxidant status. *Phytotherapy Research*, 25(11), pp.1618–1622.

Stapley, J. et al., 2010. Adaptation genomics: the next generation. *Trends in Ecology & Evolution*, 25(12), pp.705–712.

Steege, H. et al., 2015. Estimating the global conservation status of more than 15,000 Amazonian tree species. *Sci. Adv*, 1(e), pp.9–11.

Stefanowska, M., Kuraś, M. & Kacperska, A., 2002. Low temperature-induced

modifications in cell ultrastructure and localisation of phenolics in winter oilseed rape (*Brassica napus* L. v*ar. oleifer*a L.) leaves. *Annals of Botany*, 90(5), pp.637–645.

Stewart, C.N. & Via, L.E., 1993. A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. *BioTechniques*, 14(5), pp.748–50.

Stock, D. et al., 2000. The rotary mechanism of ATP-synthase. *Current Opinion in Structural Biology*, 10(6), pp.672–679.

Stoger, E. et al., 2014. Plant molecular pharming for the treatment of chronic and infectious diseases. *Annual Review of Plant Biology*, 65(1), pp.743–768.

Stone, S.E. et al., 2017. Cell-selective proteomics for biological discovery. *Current Opinion in Chemical Biology*, 36, pp.50–57.

Stratil, P., Klejdus, B. & Kuban, V., 2006. Determination of total content of phenolic compounds and their antioxidant activity in vegetables: Evaluation of spectrophotometric methods. *Journal of Agricultural and Food Chemistry*, 54, pp.607–616.

Stuart L. Pimm et al., 1995. The future of biodiversity. *Science*, 269(5222), p.347.

Su, J. et al., 2018. Constant vigilance: plant functions guarded by resistance proteins. *Plant Journal*, 93(4), pp.637–650.

Su, S. & Higashiyama, T., 2018. Arabinogalactan proteins and their sugar chains: functions in plant reproduction, research methods, and biosynthesis. *Plant Reproduction*, 31(1), pp.67–75.

Sun, H. et al., 2016. Physiological and proteomic analysis of selenium-mediated tolerance to Cd stress in cucumber (*Cucumis sativus* L.). *Ecotoxicology and Environmental Safety*, 133, pp.114–126.

Sun, H. et al., 2016. Proteomic analysis of amino acid metabolism differences between wild and cultivated Panax ginseng. *Journal of Ginseng Research*, 40(2), pp.113–120.

Sun, Q. et al., 2016. A label-free differential proteomics analysis reveals the effect of melatonin on promoting fruit ripening and anthocyanin accumulation upon postharvest in tomato. *Journal of Pineal Research*, pp.138–153.

Sundström, J.F. et al., 2009. Tudor staphylococcal nuclease is an evolutionarily conserved component of the programmed cell death degradome. *Nature Cell Biology*, 11(11), pp.1347–1354.

Sung, T. Van et al., 2008. Isolation and structural characterisation of phenolic glycoside and triterpenes in *Celastrus hindsii* Benth. *Journal of Chemistry*, 46(February), pp.224–228.

Suzuki, N. et al., 2014. Abiotic and biotic stress combinations. *New Phytologist*, 203(1), pp.32–43.

Swain, T. & Hillis, W.E., 1959. The phenolic constituents of *Prunus domestica*. I.The quantitative analysis of phenolic constituents. *Journal of the Science of Food and Agriculture*, 10(1), pp.63–68.

Syamkumar, S., Lowarence, B. & Sasikumar, B., 2003. Isolation and amplification of DNA from rhizomes of turmeric and ginger. *Plant Molecular Biology Reporter*, 21(2), pp.171-171.

Szabados, L. & Savouré, A., 2010. Proline: a multifunctional amino acid. *Trends in Plant Science*, 15(2), pp.89–97.

Szydłowska-Czerniak, A. & Szłyk, E., 2003. Spectrophotometric determination of total phosphorus in rape seeds and oils at various stages of technological process:

Calculation of phospholipids and non-hydratable phospholipids contents in rapeseed oil. *Food Chemistry*, 81(4), pp.613–619.

Tabassum, A., Bristow, R.G. & Venkateswaran, V., 2010. Ingestion of selenium and other antioxidants during prostate cancer radiotherapy: a good thing? *Cancer Treatment Reviews*, 36(3), pp.230–4.

Tahira, R. & Saleem, M., 2013. Evaluation of genetic diversity of raya (*Brassica juncea*) through RAPD markers. *International Journal of Agriculture and Biology*, 15(6), pp.1163–1168.

Tai, T.H. & Tanksley, S.D., 1990. A rapid and inexpensive method for isolation of total DNA from dehydrated plant tissue. *Plant Molecular Biology Reporter*, 8(4), pp.297–303.

Tajima, T. et al., 2011. Biochemical and molecular characterisation of senescence-related cysteine protease-cystatin complex from spinach leaf. *Physiologia Plantarum*, 141(2), pp.97–116.

Takahashi, Y. et al., 2015. Current pharmacological therapies for nonalcoholic fatty liver disease/nonalcoholic steatohepatitis. *World Journal of Gastroenterology*, 21(13), pp.3777–3785.

Tan, D., Thu, P. & Dell, B., 2012. Invasive plant species in the National Parks of Vietnam. *Forests*, 3(4), pp.997–1016.

Tan, E.C. et al., 2012. Proteomic analysis of cell suspension cultures of *Boesenbergia rotunda* induced by phenylalanine: Identification of proteins involved in flavonoid and phenylpropanoid biosynthesis pathways. *Plant Cell, Tissue and Organ Culture*, 111(2), pp.219–229.

Tan, L. et al., 2012. Arabinogalactan-proteins and the research challenges for these enigmatic plant cell surface proteoglycans. *Frontiers in Plant Science*, 3(June), pp.1–10.

Tan, S.C. & Yiap, B.C., 2009. DNA, RNA, and protein extraction: the past and the present. *Journal of Biomedicine & Biotechnology*, 2009, p.574398.

Tanaka, R. et al., 1999. Reduced activity of geranylgeranyl reductase leads to loss of chlorophyll and tocopherol and to partially geranylgeranylated chlorophyll in transgenic tobacco plants expressing antisense RNA for geranylgeranyl reductase. *Plant Physiol*, 120, pp.695– 704.

Tattini, M. et al., 2004. Differential accumulation of flavonoids and hydroxycinnamates in leaves of *Ligustrum vulgare* under excess light and drought stress. *New Phytologist*, 163(3), pp.547–561.

Tausz, M., Sircelj, H. & Grill, D., 2004. The glutathione system as a stress marker in plant ecophysiology: Is a stress-response concept valid? *Journal of Experimental Botany*, 55(404), pp.1955–1962.

Taylor, N.L. et al., 2014. Selected reaction monitoring to determine protein abundance in *Arabidopsis* using the *Arabidopsis* proteotypic predictor. *Plant Physiology*, 164(2), pp.525–36.

Te-Chato, S., 2000. Random amplified polymorphic DNA (RAPD) markers for genetic analysis in somaclones of mangosteen (*Garcinia mangostana* L.). *Thai Journal of Agricultural Science*, 33(3/4), pp.137–145.

Terpinc, P. et al., 2011. Antioxidant properties of 4-vinyl derivatives of hydroxycinnamic acids. *Food Chemistry*, 128(1), pp.62–69.

Testerink, C. & Munnik, T., 2011. Molecular, cellular, and physiological responses to phosphatidic acid formation in plants. *Journal of Experimental Botany*, 62(7),

pp.2349–2361.

Testerink, C. & Munnik, T., 2005. Phosphatidic acid: A multifunctional stress signalling lipid in plants. *Trends in Plant Science*, 10(8), pp.368–375.

Thai, H.T., Lee, D.K. & Woo, S.Y., 2010. Growth of several indigenous species in the degraded forest in northern Vietnam. *International Journal of the Physical Sciences*, 5(17), pp.2664–2671.

Thajuddin, N. et al., 2010. Morphological and genetic diversity of symbiotic cyanobacteria from cycads. *Journal of Basic Microbiology*, 50(3), pp.254–265.

The Asian Foundation, 2012. *Medicinal plant index of the Daos in Ba Vi*, Hanoi: Cong ty Truyen thong ICON. In Vietnamese.

Thomas, S.M., Moreno, R.F. & Tilzer, L.L., 1989. DNA extraction with organic solvents in gel barrier tubes. *Nucleic Acids Research*, 17(13), p.5411.

Thuy, T.T., Cuong, N.H. & Sung, T. Van, 2007. Triterpenes from *Celastrus hindsii* Benth. *Journal of Chemistry*, pp.373–376.

Tiago, O. et al., 2017. Plant secondary metabolites and its dynamical systems of induction in response to environmental factors: A review. *African Journal of Agricultural Research*, 12(2), pp.71–84.

Tianniam, S., Bamba, T. & Fukusaki, E., 2010. Pyrolysis GC-MS-based metabolite fingerprinting for quality evaluation of commercial *Angelica acutiloba* roots. *JBIOSC*, 109(1), pp.89–93.

Tietze, F., 1969. Enzymic method for quantitative determination of nanogram amounts of total and oxidized glutathione: applications to mammalian blood and other tissues. *Analytical Biochemistry*, 27(3), pp.502–522.

Timabud, T., Sanitchon, J. & Pongdontri, P., 2013. A modified ferrous oxidation-xylenol orange assay for lipoxygenase activity in rice grains. *Food Chemistry*, 141(3), pp.2405–2411.

Tiwari, K.L., Jadhav, S.K. & Gupta, S., 2012. Modified CTAB technique for isolation of DNA from some medicinal plants. *Research Journal of Medicinal Plants*, 6(1), pp.65–73.

Toth, G.B. & Pavia, H., 2001. Removal of dissolved brown algal phlorotannins using insoluble polyvinylpolypyrrolidone (PVPP). *Journal of Chemical Ecology*, 27(9), pp.1899–910.

Traber, M.G. & Atkinson, J., 2007. Vitamin E , antioxidant and nothing more. *Free Radical Biology & Medicine*, 43, pp.4–15.

Tran, V.C., 2001. *Additional research to selection of main tree species using for production forest plantation in Central Highland. Scientific and technological research results in forestry from 1996 to 2000*, Hanoi: Agriculture publishing House.

Traylor-Knowles, N. et al., 2017. Early transcriptional responses during heat stress in the coral *Acropora hyacinthus*. *The Biological Bulletin*, 232(2), pp.91–100.

Treutter, D., 2006. Significance of flavonoids in plant resistance: a review. *Environ Chem Lett*, 4, pp.147–157.

Tripathi, N. et al., 2012. Lack of genetic diversity of a critically endangered important medicinal plant *Chlorophytum borivilianum* in Central India revealed by AFLP markers. *Physiology and Molecular Biology of Plants,* 18(2), pp.161–167.

Trotel-Aziz, P. et al., 2003. The control of proline consumption by abscisic acid during osmotic stress recovery of canola leaf discs. *Physiologia Plantarum*, 117, pp.213–221.

Tsikas, D., 2017. Assessment of lipid peroxidation by measuring malondialdehyde

(MDA) and relatives in biological samples : Analytical and biological challenges. *Analytical Biochemistry*, 524, pp.13–30.

Tsikas, D. et al., 1999. Determination of S-Nitrosoglutathione in human and rat plasma by high-performance liquid chromatography with fluorescence and ultraviolet absorbance detection after pre-column derivatization with o-phthalaldehyde. *Analytical Biochemistry*, 273(1), pp.32–40.

Tsuchiya, N. et al., 2007. SND1, a component of RNA-induced silencing complex, is up-regulated in human colon cancers and implicated in early-stage colon carcinogenesis. *Cancer Research*, 67(19), pp.9568–9576.

Tsukuda, S. et al., 2006. Characterization of cDNAs encoding two distinct miraculin-like proteins and stress-related modulation of the corresponding mRNAs in *Citrus jambhiri* Lush. *Plant Molecular Biology*, 60(1), pp.125–136.

Tuna, A.L. et al., 2008. The combined effects of gibberellic acid and salinity on some antioxidant enzyme activities, plant growth parameters and nutritional status in maize plants. *Environmental and Experimental Botany*, 62(1), pp.1–9.

Tworak, A. et al., 2016. Six *Medicago truncatula* Dicer-like protein genes are expressed in plant cells and upregulated in nodules. *Plant Cell Reports*, 35(5), pp.1043–1052.

Tyagi, T. & Agarwal, M., 2017. Antioxidant properties and phenolic compounds in methanolic extracts of *Eichhornia crassipes*. *Phytochemistry,* 11(2), pp.85-89.

Unhelkar, M.H. et al., 2017. Structure prediction and network analysis of chitinases from the Cape sundew, *Drosera capensis*. *Biochimica et Biophysica Acta - General Subjects*, 1861(3), pp.636–643.

Upendar, J., 2017. Protein sequence design and analysis: BLAST-based computational approaches. *International Journal of Electronics, Electrical and Computational System*, 6(11), pp.379–382.

Uprety, Y. et al., 2012. Traditional use of medicinal plants in the boreal forest of Canada: review and perspectives. *Journal of Ethnobiology and Ethnomedicine*, 8(7), pp.1-14.

Vâlcu, C.-M. & Schlink, K., 2006. Reduction of proteins during sample preparation and two-dimensional gel electrophoresis of woody plant samples. *Proteomics*, 6(5), pp.1599–605.

Valenzuela, A., 1991. The biological significance of malondialdehyde determination in the assessment of oxidative tissue stress. *Life Sciences*, 48(4), pp.301–309.

Van Cutsem, E. et al., 2011. Gel-based and gel-free proteomic analysis of *Nicotiana tabacum* trichomes identifies proteins involved in secondary metabolism and in the (a)biotic stress response. *Proteomics*, 11(3), pp.440–454.

van der Hoorn, R.A.L., 2008. Plant proteases: from phenotypes to molecular mechanisms. *Annual Review of Plant Biology*, 59, pp.191–223.

Van Ness, L.K. et al., 2016. Mass spectrometric-based selected reaction monitoring of protein phosphorylation during symbiotic signalling in the model legume, *Medicago truncatula*. *PLoS ONE*, 11(5), pp.1–15.

Vanderschuren, H. et al., 2013. Proteomics of model and crop plant species: Status, current limitations and strategic advances for crop improvement. *Journal of Proteomics*, 93, pp.5–19.

Vanderweyde, T. et al., 2013. Role of stress granules and RNA-binding proteins in neurodegeneration: A mini review. *Gerontology*, 59(6), pp.524–533.

Varma, A., Padh, H. & Shrivastava, N., 2007. Plant genomic DNA isolation: an art or a science. *Biotechnology Journal*, 2(3), pp.386–92.

Ve, T., Williams, S.J. & Kobe, B., 2015. Structure and function of Toll/interleukin-1 receptor/resistance protein (TIR) domains. *Apoptosis*, 20(2), pp.250–261.

Vensel, W.H., Tanaka, C.K. & Altenbach, S.B., 2014. Protein composition of wheat gluten polymer fractions determined by quantitative two-dimensional gel electrophoresis and tandem mass spectrometry. *Proteome Science*, 12(8), pp.1-13.

Verpoorte, R., Contin, A. & Memelink, J., 2002. Biotechnology for the production of secondary plant metabolites. *Phytochemistry Reviews*, 1(1), pp.13–25.

Vieira, R.F. et al., 2001. Genetic diversity of *Ocimum gratissimum* L. based on volatile oil constituents, flavonoids and RAPD markers. *Biochemical Systematics and Ecology*, 29(3), pp.287–304.

Virk, P.S. et al., 1995. The identification of duplicate accessions with a rice germplasm collection using RAPD analysis. *Theoretical Applied Genetics*, 90, pp.1049–1055.

Volis, S. & Blecher, M., 2010. *Quasi in situ*: A bridge between *ex situ* and *in situ* conservation of plants. *Biodiversity and Conservation*, 19(9), pp.2441–2454.

Voronova, A. et al., 2014. Stress-induced transcriptional activation of retrotransposon-like sequences in the Scots pine (*Pinus sylvestris* L.) genome. *Tree Genetics and Genomes*, 10(4), pp.937–951.

Walker, M.A. & McKersie, B.D., 1993. Role of the ascorbate-glutathione antioxidant system in chilling resistance of tomato. *J. Plant Physiol.*, 141, pp.234–239.

Waller, D.M., 2015. Genetic rescue: a safe or risky bet? *Molecular Ecology*, 24(11), pp.2595–2597.

Walters, T.W. & Decker-Walters, D.S., 1991. Patterns of allozyme diversity in the west indies cycad *Zamia pumila* (Zamiaceae). *American Journal of Botany*, 78(3), pp.436–445.

Wang, C. et al., 2008. The cytoplasmic-localised, cytoskeletal-associated RNA binding protein OsTudor-SN: Evidence for an essential role in storage protein RNA transport and localisation. *Plant Journal*, 55(3), pp.443–454.

Wang, H. & Wang, K.W., 2016. Lupene-type triterpenoids from *Celastrus oblanceifolius. Chemistry of Natural Compounds*, 52(4), pp.764–765.

Wang, L. et al., 2015. Proteomic analysis of changes in the *Kandelia candel* chloroplast proteins reveals pathways associated with salt tolerance. *Plant Science*, 231, pp.159–172.

Wang, L. et al., 2014. Proteomic analysis of salt-responsive proteins in the leaves of mangrove *Kandelia candel* during short-term stress. *PLoS ONE*, 9(1), p.e83141.

Wang, M. et al., 2013. Quantitative proteomic analysis reveals the mode-of-action for chronic mercury hepatotoxicity to marine medaka (*Oryzias melastigma*). *Aquatic Toxicology*, 130–131, pp.123–131.

Wang, N. et al., 2012. Prognostic impact of metadherin–SND1 interaction in colon cancer. *Molecular Biology Reports*, 39(12), pp.10497–10504.

Wang, P. et al., 2018. Evolutionary and functional characterisation of leucoanthocyanidin reductases from *Camellia sinensis*. *Planta*, 247(1), pp.139–154.

Wang, W. et al., 2006. A universal and rapid protocol for protein extraction from recalcitrant plant tissues for proteomic analysis. *Electrophoresis*, 27(13), pp.2782–2786.

Wang, W. et al., 2003. Protein extraction for two-dimensional electrophoresis from olive leaf, a plant tissue containing high levels of interfering compounds. *Electrophoresis*, 24(14), pp.2369–2375.

Wang, W., Tai, F. & Chen, S., 2008. Optimizing protein extraction from plant tissues for

enhanced proteomics analysis. *Journal of Separation Science*, 31(11), pp.2032–209.

Wang, X. et al., 2007. A protein extraction method compatible with proteomic analysis for the euhalophyte *Salicornia europaea. Electrophoresis*, 28(21), pp.3976–3987.

Wang, X. et al., 2013. Comparative proteomics of *Thellungiella halophila* leaves from plants subjected to salinity reveals the importance of chloroplastic starch and soluble sugars in halophyte salt tolerance. *Molecular & Cellular Proteomics*, 12(8), pp.2174–2195.

Wang, X., Wang, Z. & Zou, Y., 1996. An improved procedure for the isolation of nuclear DNA from leaves of wild grapevine dried with silica gel. *Plant Molecular Biology Reporter*, 14(4), pp.369–373.

Wang, Y. et al., 2016. Comparative proteomic analysis of the response to silver ions and yeast extract in *Salvia miltiorrhiza* hairy root cultures. *Plant Physiology and Biochemistry*, 107, pp.364–373.

Wang, Z. et al., 2011. Probing paramagnetic species in titania-based heterogeneous photocatalysis by electron spin resonance (ESR) spectroscopy - A mini review. *Chemical Engineering Journal*, 170(2–3), pp.353–362.

Warude, D. et al., 2003. DNA isolation from fresh, dry plant samples with highly acidic tissue extracts DNA isolation from fresh and dry plant samples. *Plant Molecular Biology Reporter*, 21, p.467a–467f.

Wayne, R.K. & Bradley, S.H., 2016. Hybridization and endangered species protection in the molecular era. *Molecular Ecology*, 25(11), pp.2680–2689.

Webster, G.R., Teh, A.Y.H. & Ma, J.K.C., 2017. Synthetic gene design: The rationale for codon optimisation and implications for molecular pharming in plants. *Biotechnology and Bioengineering*, 114(3), pp.492–502.

Wee, A.K., Surget-Groba, Y. & Corlett, R., 2015. Genetic optimisation of trees. *BG Journal*, 12(2), pp.18–20.

Wei, C.L. et al., 2016. An improved DNA marker technique for genetic characterisation using RAMP-PCR with high-GC primers. *Genetics and Molecular Research*, 15(3), pp.1-12.

Welch, A. a & Hardcastle, A.C., 2014. The effects of flavonoids on bone. *Current Osteoporosis Reports*, 12(2), pp.205–210.

Welsh, J. & Mcclelland, M., 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acid Research*, 18(24), pp.6531–6535.

Welti, R. et al., 2002. Profiling membrane lipids in plant stress responses. *The Journal of Biological Chemistry*, 277(35), pp.31994–32002.

Wen, J., Li, J. & Walker, J.C., 2012. Overexpression of a serine carboxypeptidase increases carpel number and seed production in *Arabidopsis thaliana*. *Food and Energy Security*, 1(1), pp.61–69.

Wen, P.F. et al., 2015. Accumulation of flavanols and expression of leucoanthocyanidin reductase induced by postharvest UV-C irradiation in grape berry. *Genetics and Molecular Research*, 14(3), pp.7687–7695.

Weng, J.K., 2014. The evolutionary paths towards complexity: A metabolic perspective. *New Phytologist*, 201(4), pp.1141–1149.

Westermeier, R., 2014. Looking at proteins from two dimensions: a review on five decades of 2D electrophoresis. *Archives of Physiology and Biochemistry*, 3455, pp.1–5.

Wheeler, N. & Sederoff, R., 2008. Role of genomics in the potential restoration of the American chestnut. *Tree Genetics & Genomes*, 5, pp.181-187.

Widodo et al., 2009. Metabolic responses to salt stress of barley (*Hordeum vulgare* L.) cultivars, Sahara and Clipper, which differ in salinity tolerance. *Journal of Experimental Botany*, 60(14), pp.4089–4103.

Wiernsperger, N.F., 2003. Oxidative stress as a therapeutic target in diabetes: revisiting the controversy. *Diabetes & Metabolism*, 29(6), pp.579–85.

Wilcock, C. & Neiland, R., 2002. Pollination failure in plants: why it happens and when it matters. *Trends in Plant Science*, 7(6), pp.270–277.

Wilkins, M.R. et al., 1998. Multiple parameter cross-species protein identification using MultiIdent: A world-wide web-accessible tool. *Electrophoresis*, 19(18), pp.3199–3206.

Wilkins, M.R. et al., 1996. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnology & Genetic Engineering Reviews*, 13, pp.19–50.

Wilkins, M.R. & Williams, K.L., 1997. Cross-species protein identification using amino acid composition, peptide mass fingerprinting, isoelectric point and molecular mass: a theoretical evaluation. *Journal of Theoretical Biology*, 186(1), pp.7–15.

Williams, C.E. & Ronald, P.C., 1994. PCR template-DNA isolated quickly from monocot and dicot leaves without tissue homogenization. *Nucleic Acids Research*, 22(10), pp.1917–8.

Williams, J.G. et al., 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*, 18(22), pp.6531–5.

Williams, J.G.K. et al., 1993. Genetic analysis using random amplified polymorphic DNA markers. In R. Wu, ed. *Methods in Enzymology*. Academic Press, pp. 704–740.

Winkel-Shirley, B., 2002. Biosynthesis of flavonoids and effects of stress. *Current Opinion in Plant Biology*, 5(3), pp.218–23.

Winterbourn, C.C., 2016. Revisiting the reactions of superoxide with glutathione and other thiols. *Archives of Biochemistry and Biophysics*, 595, pp.68–71.

Witters, E. et al., 2003. Fast liquid chromatography coupled to electrospray tandem mass spectrometry peptide sequencing for cross-species protein identification. *Rapid Communications in Mass Spectrometry*, 17(19), pp.2188–2194.

Witwer, K.W. & Hirschi, K.D., 2014. Transfer and functional consequences of dietary microRNAs invertebrates: Concepts in search of corroboration. *BioEssays*, 36(4), pp.394–406.

Woerdenbag, H.J. et al., 2012. Vietnamese traditional medicine from a pharmacist's perspective. *Expert Review of Clinical Pharmacology*, 5(4), pp.459–77.

Wong, K.L. et al., 2014. Bioactive proteins and peptides isolated from Chinese medicines with pharmaceutical potential. *Chinese Medicine*, 9(1), p.19.

Wootton-Beard, P.C., Moran, A. & Ryan, L., 2011. Stability of the total antioxidant capacity and total polyphenol content of 23 commercially available vegetable juices before and after *in vitro* digestion measured by FRAP, DPPH, ABTS and Folin–Ciocalteu methods. *Food Research International*, 44(1), pp.217–224.

Wright, J.C., Beynon, R.J. & Hubbard, S.J., 2010. Cross-species proteomics. In S. J. Hubbard & A. R. Jones, eds. *Proteome Bioinformatics*. Totowa, NJ: Humana Press, pp. 123–135.

Wu, K.S. & Tanksley, S.D., 1993. Abundance, polymorphism and genetic mapping of microsatellites in rice. *Molecular & General Genetics*, 241(1–2), pp.225–35.

Wu, S.R. et al., 2017. The Kunitz domain i of hepatocyte growth factor activator inhibitor-2 inhibits matriptase activity and invasive ability of human prostate cancer cells.

Scientific Reports, 7(1), pp.1–19.

Wu, X., Gong, F. & Wang, W., 2014. Protein extraction from plant tissues for 2DE and its application in proteomic analysis. *Proteomics*, 14(6), pp.645–58.

Wujeska-Klause, A., Bossinger, G. & Tausz, M., 2015. Responses to heatwaves of gas exchange, chlorophyll fluorescence and antioxidants ascorbic acid and glutathione in congeneric pairs of Acacia and Eucalyptus species from relatively cooler and warmer climates. *Trees - Structure and Function*, 29(6), pp.1929–1941.

Xiao, J., Chen, T. & Cao, H., 2014. Flavonoid glycosylation and biological benefits. *Biotechnology Advances*.

Xiao, L. et al., 2004. ISSR variation in the endemic and endangered plant *Cycas guizhouensis* (Cycadaceae). *Annals of Botany*, 94(1), pp.133–8.

Xiao, L.Q. & Gong, X., 2006. Genetic differentiation and relationships of populations in the *Cycas balansae* complex (Cycadaceae) and its conservation implications. *Annals of Botany*, 97(5), pp.807–812.

Xing, A., Pan, L. & Gao, J., 2018. p100 functions as a metastasis activator and is targeted by tumour suppressing microRNA-320a in lung cancer. *Thoracic Cancer*, 9(1), pp.152–158.

Xing, H. et al., 2018. Genome-wide investigation of pentatricopeptide repeat gene family in poplar and their expression analysis in response to biotic and abiotic stresses. *Scientific Reports*, 8(1), pp.1–9.

Xu, J. & Chua, N.H., 2011. Processing bodies and plant development. *Current Opinion in Plant Biology*, 14(1), pp.88–93.

Xue, H., Chen, X. & Li, G., 2007. Involvement of phospholipid signalling in plant growth and hormone effects. *Current Opinion in Plant Biology*, 10(5), pp.483–489.

Yahiaoui, N. et al., 2004. Genome analysis at different ploidy levels allows cloning of the powdery mildew resistance gene Pm3b from hexaploid wheat. *Plant Journal*, 37(4), pp.528–538.

Yan, C. et al., 2014. Tudor-SN, a component of stress granules, regulates growth under salt stress by modulating GA20ox3 mRNA levels in *Arabidopsis. Journal of Experimental Botany*, 65(20), pp.5933–5944.

Yang, H. et al., 2012. Application of next-generation sequencing for rapid marker development in molecular plant breeding : a case study on anthracnose disease resistance in *Lupinus angustifolius* L. *BMC Genomics*, 13(318), pp 1-11.

Yang, M. et al., 2009. Phytochemical analysis of traditional Chinese medicine using liquid chromatography coupled with mass spectrometry. *Journal of Chromatography,* 1216(11), pp.2045–2062.

Yang, S.L. & Meerow, A.W., 1996. The *Cycas pectinata* (Cycadaceae) complex: Genetic structure and gene flow. *International Journal of Plant Sciences*, 157(4), pp.468–483.

Yang, S.L. & Meerow, A.W., 1999. Genetic variation in Chinese cycad population. In *Biology and conservation of cycads. Proceedings of the Fourth International Conference on Cycad Biology. Beijing: International Academic Publishers*. pp. 175–186.

Yang, W. et al., 1996. Comparison of DNA marker technologies in characterizing plant genome diversity: variability in Chinese sorghums. *Crop Science*, 36, pp.1669–1676.

Yang, W. et al., 2015. The K-segments of wheat dehydrin WZY2 are essential for its protective functions under temperature stress. *Frontiers in Plant Science*, 6(June),

pp.1–9.

Yang, Y. et al., 2008. Isolation and characterization of microsatellite markers for *Cycas debaoensis* Y. C. Zhong et C. J. Chen (Cycadaceae). *Molecular Ecology Resources*, 8(4), pp.913–915.

Yao-Haur Kuo, Chen, C.F. & Kuo, L.M.Y., 1995. Celahinine A, a new sesquiterpene pyridine alkaloid from *Celastrus hindsii*. *Journal of Natural Products*, 58(11), pp.1735–1738.

Yao, Y., Yang, Y.W. & Liu, J.Y., 2006. An efficient protein preparation for proteomic analysis of developing cotton fibers by 2-DE. *Electrophoresis*, 27(22), pp.4559–4569.

Yasmin, T. et al., 2015. A germin-like protein gene of rice increased superoxide dismutase activity in transformed tobacco. *Biologia Plantarum*, 59(3), pp.456–462.

Yeh, F., Yang, R. & Boyle, T., 1999. *Popgene 1.31, the user-friendly shareware for population genetic analysis*, Molecular Biology and Biotechnology Centre, University of Alberta, Edmonton.

Yeh, S. et al., 2000. Chitinase genes responsive to cold encode antifreeze proteins in winter cereals. *Plant Physiology*, 124(3), pp.1251–64.

Yeo, S.C., Cheung, C.K. & Barratt, J., 2017. New insights into the pathogenesis of IgA nephropathy. *Pediatric Nephrology*, 33(5), pp.763–777.

Ying, Y.-M. et al., 2017. Chemical constituents of *Celastrus rugosus*. *Chemistry of Natural Compounds*, 53(3), pp.589–591.

Yoo, B.K. et al., 2011. Increased RNA-induced silencing complex (RISC) activity contributes to hepatocellular carcinoma. *Hepatology*, 53(5), pp.1538–1548.

Yoon, H.J. et al., 2000. Induced accumulation of triterpenoids in *Scutellaria baicalensis* suspension cultures using a yeast elicitor. *Biotechnology Letters*, 22(13), pp.1071–1075.

Yordanov, I., Velikova, V. & Tsonev, T., 2000. Plant responses to drought, acclimation, and stress tolerance. *Photosynthetica*, 38(2), pp.171–186.

Younus, M., 2015. Ethno botanical study and traditional uses of *Celastrus paniculatus*. *International Journal of Innovative Science, Engineering & Technology*, 2(11), pp.139–143.

Yu, L. et al., 2010. Phosphatidic acid mediates salt stress response by regulation of MPK6 in *Arabidopsis thaliana*. *New Phytologist*, 188(3), pp.762–773.

Yu, L. et al., 2017. SND1 acts as a novel gene transcription activator recognizing the conserved motif domains of smad promoters, inducing TGFβ1 response and breast cancer metastasis. *Oncogene*, 36(27), pp.3903–3914.

Yu, Z.X. et al., 2012. The jasmonate-responsive AP2/ERF transcription factors AaERF1 and AaERF2 positively regulate artemisinin biosynthesis in *Artemisia annua* L. *Molecular Plant*, 5(2), pp.353–365.

Yuan, Y. et al., 2012. Genetic variation and metabolic pathway intricacy govern the active compound content and quality of the Chinese medicinal plant *Lonicera japonica* Thunb. *BMC Genomics*, 13(195), pp.1-17.

Zahid, A. et al., 2017. Arabinogalactan proteins from Baobab and Acacia seeds influence innate immunity of human keratinocytes *in vitro*. *Journal of Cellular Physiology*, 232(9), pp.2558–2568.

Zamora-Ros, R. et al., 2016. Dietary polyphenol intake in Europe: the European prospective investigation into cancer and nutrition (EPIC) study. *European Journal of Nutrition*, 55(4), pp.1359–1375.

Zandalinas, S.I. et al., 2017. Activation of secondary metabolism in *Citrus* plants is associated to sensitivity to combined drought and high temperatures. *Frontiers in Plant Science*, 7(January), pp.1–17.

Zeng, Y. et al., 2011. A proteomic analysis of the chromoplasts isolated from sweet orange fruits *Citrus sinensis* (L.) Osbeck]. *Journal of Experimental Botany*, 62(15), pp.5297–5309.

Zhang, H. & Tsao, R., 2016. Dietary polyphenols, oxidative stress and antioxidant and anti-inflammatory effects. *Current Opinion in Food Science*, 8, pp.33–42.

Zhang, J. et al., 2017. PDM3, a pentatricopeptide repeat-containing protein, affects chloroplast development. *Journal of Experimental Botany*, 267(20), pp.27–37.

Zhang, K. et al., 2014. Arabidopsis ABCG14 protein controls the acropetal translocation of root-synthesized cytokinins. *Nature Communications*, 5, p.3274.

Zhang, L. et al., 2012. Exogenous plant MIR168a specifically targets mammalian LDLRAP1: evidence of cross-kingdom regulation by microRNA. *Cell Research*, 22, p.107.

Zhang, L. et al., 2014. Proteomics analysis of *Mahonia bealei* leaves with induction of alkaloids via combinatorial peptide ligand libraries. *Journal of Proteomics*, 110, pp.59–71.

Zhang, L. et al., 2005. The genetic diversity and similarities among *Kengyilia* species based on random amplified microsatellite polymorphism (RAMP). *Genetic Resources and Crop Evolution*, 52(8), pp.1011–1017.

Zhang, M. et al., 2009. Chemical and pharmacological studies of the plants from genus *Celastrus*. *Chemistry and Biodiversity*, 6, pp.146–161.

Zhang, X. et al., 2017. A proteolytic regulator controlling chalcone synthase stability and flavonoid biosynthesis in *Arabidopsis. The Plant Cell*, 29(5), pp.1157–1174.

Zhang, X. et al., 2015. Down-regulation of Kelch domain-containing F-box protein in *Arabidopsis* enhances the production of (Poly)phenols and tolerance to ultraviolet radiation. *Plant Physiology*, 167(2), pp.337–350.

Zhang, X. et al., 2017. Multiple functional self-association interfaces in plant TIR domains. *Proceedings of the National Academy of Sciences*, 114(10), pp.E2046–E2052.

Zhang, X., Gou, M. & Liu, C.J., 2013. *Arabidopsis* Kelch repeat F-box proteins regulate phenylpropanoid biosynthesis via controlling the turnover of phenylalanine ammonia-lyase. *The Plant Cell*, 25(12), pp.4994–5010.

Zhang, X. & Liu, C.J., 2015. Multifaceted regulations of gateway enzyme phenylalanine ammonia-lyase in the biosynthesis of phenylpropanoids. *Molecular Plant*, 8(1), pp.17–27.

Zhang, Y. et al., 2004. Expression of RPS4 in tobacco induces an AvrRps4-independent HR that requires EDS1, SGT1 and HSP90. *Plant Journal*, 40(2), pp.213–224.

Zhang, Y. et al., 2018. Overexpression of germin-like protein GmGLP10 enhances resistance to Sclerotinia sclerotiorum in transgenic tobacco. *Biochemical and Biophysical Research Communications*, 497(1), pp.160–166.

Zhang, Z.Y., Chen, Y.Y. & Li, D.Z., 2005. Detection of low genetic variation in a critically endangered Chinese pine, *Pinus squamata*, using RAPD and ISSR markers. *Biochemical Genetics*, 43(5–6), pp.239–249.

Zhao, H. et al., 2016. An analysis of genetic diversity in *Marphysa sanguinea* from different geographic populations using ISSR polymorphisms. *Biochemical Systematics and Ecology*, 64, pp.65–69.

Zhao, J. et al., 2007. Analysis of gene expression profiles in response to *Sclerotinia sclerotiorum* in *Brassica napus*. *Planta*, 227(1), pp.13–24.

Zhao, L. et al., 2013. Assessment of the genetic diversity and genetic relationships of pomegranate (*Punica granatum* L.) in China using RAMP markers. *Scientia Horticulturae*, 151, pp.63–67.

Zhao, X. et al., 2014. Comparative proteomics of *Euphorbia kansui* Liou milky sap at two different developmental stages. *Plant Physiology and Biochemistry*, 79, pp.60–65.

Zhao, Y. et al., 2017. A microRNA cluster (let-7c, miRNA-99a, miRNA-125b, miRNA-155 and miRNA-802) encoded at chr21q21.1-chr21q21.3 and the phenotypic diversity of Down's syndrome (DS; trisomy 21). *Journal of Nature and Science*, 3(9), p.e446.

Zhao, Y., Cong, L. & Lukiw, W.J., 2018. Plant and animal microRNAs (miRNAs) and their potential for inter-kingdom communication. *Cellular and Molecular Neurobiology*, 38(1), pp.133–140.

Zhen, Y. & Shi, J., 2011. Evaluation of sample extraction methods for proteomic analysis of coniferous seeds. *Acta Physiologiae Plantarum*, 33(5), pp.1623–1630.

Zhixiang, Z. & Funston, A.M., 2008. *Celastrus* Linnaeus, Sp. Pl. 1: 196. 1753, nom. cons. *Fl. China*, 11, pp.466–474.

Zhong, R. et al., 2002. Mutation of a chitinase-like gene causes ectopic deposition of lignin, aberrant cell shapes, and overproduction of ethylene. *The Plant Cell Online*, 14(1), pp.165–179.

Zhu, L. et al., 2013. Characterization of Tudor-staphylococcal nuclease-containing granules in the silkworm, *Bombyx mori*. *Insect Biochemistry and Molecular Biology*, 43(8), pp.664–674.

Zhu, W. & Tan, S., 2017. Tudor-SN protein expression in colorectal cancer and its association with clinical characteristics. *Open Life Science*s, 12(1), pp.237–242.

Zlatev, Z.S. et al., 2006. Comparison of resistance to drought of three bean cultivars. *Biologia Plantarum*, 50(3), pp.389–394.

Zsigmond, L. et al., 2008. *Arabidopsis* PPR40 connects abiotic stress responses to mitochondrial electron transport. *Plant Physiology*, 146(4), pp.1721–1737.

Zubarev, R.A., 2013. The challenge of the proteome dynamic range and its implications for in-depth proteomics. *Proteomics*, 13(5), pp.723–726.

Zulak, K.G., Khan, M.F., et al., 2009. Plant defence responses in opium poppy cell cultures revealed by liquid chromatography-tandem mass spectrometry proteomics. *Molecular & Cellular Proteomics*, 8(1), pp.86–98.

Zulak, K.G. et al., 2009. Targeted proteomics using selected reaction monitoring reveals the induction of specific terpene synthases in a multi-level study of methyl jasmonate-treated Norway spruce (*Picea abies*). *Plant Journal*, 60(6), pp.1015–1030.

APPENDICES

Appendix 1. Supplementary genetic diversity pre-analysis results/data

Primer Selection

RAPD-PCR primers were selected at random within commercial kits where all details primers had a GC content of 50-70%. From previous studies (Sommerville 2001; De Filippis, Hoffmann & Hampp 1996; Hung, De Filippis & Buckney 2011) kits A and D from Open technologies were good at producing multiple bands in plants, but other primers were also successful.

RAMP-PCR microsatellite primers were more carefully chosen. (AT)n primers are difficult to identify and can fail to produce amplification products. (AG)n and (AC)n motifs were found to be abundant compared to (AT)n microsatellite DNA in plants. More specifically in the forest and endangered species (AG)n and (AC)n motifs were moderately abundant, but not as abundant as (GA)n microsatellites. Anchored microsatellite primes and (AC)n motifs were best with either GT or GC. Taken all together, we chose both microsatellite primers to have (AC)n motifs and one microsatellite primer having (CT)n motif with 3 different anchors CA TG, CA, respectively. RAPD and RAMP primers are given in Table 1.1 - 1.2 below.

PCR was carried out on each of the two DNA species of *C. fugax, and C. hindsii* using 10 RAPD primers of the 10 RAPD primers tested 2 were chosen for further use in RAMP. Different combinations of three microsatellite primers combined with the two chosen RAPD primers (at a ratio of 2:1 microsatellite to RAPD).

Each RAMP-PCR reaction was performed in a final volume of 25µl containing 1 x Fisher Biotech Reaction Buffer, 1M betaine, 2mM Fisher Biotech mixed dNTPs, 0.1µg/ml RNase A, 2.5 mM $MgCl_2$ (3.0 mM for RAMP-PCR); 1.0 units Taq DNA Polymerase (1.5 unit for RAMP-PCR), 25ng genomic DNA, and 12pmol of each primer. Reactions were carried out in a Biometra Personal Cycle using the following temperature profile: Denaturing at 94°C for 1 minute, then 46 cycles of denaturation at 92°C for 1 minute, annealing 35°C for 1 minute, extension 72°C for 1 minute, finally an extension at 72°C for 5 min and cooling to 4°C.

Table 1.1 Anchored microsatellite primers used to detect polymorphisms in *C. fugax,* and *C. hindsii* Bases underlined at the 5' end denote the anchors.

Primer	Sequence	Molecular weight	%GC	Tm
MSI	5' CAACACACACACAC 3'	4155	50	40.1°C
MS2	5' TGACACACACACAC 3'	4186	50	40.2°C
MS3	5' CAACTCTCTCTCTCT3'	4414	< 50	38.3°C

Table 1.2 Oligonucleotide primer sequences used to detect polymorphisms in *C. fugax* and *C. hindsii* species by RAPD analysis in Vietnam

Primer	Sequence	M.W.	% GC	Tm
Cycas fugax				
LDA-6	5’ GGTCCCTGAC 3’	3004	70	30.8°C
LDB-13	5’ TTCCCCCGCT 3’	2915	70	47.4°C
LDC-2	5’ GTGAGGCGTC 3’	3084	70	33.4°C
LDOPC-13	5’ AAGCCTCGTC 3’	2988	60	32.6°C
LDOPC-19	5’ GTTGCCAGCC 3’	3004	70	39.6°C
LDD-2	5’ GGACCCAACC 3’	2982	70	36.8°C
Celastrus hindsii				
LDA-2	5’ TGCCGAGCTG 3’	3044	70	42.4°C
LDA-7	5’ GAAACGGGTG 3’	3117	60	34.5°C
LDC-2	5’ GTGAGGCGTC 3’	3084	70	33.4°C
LDD-2	5’ GGACCCAACC 3’	2982	70	36.8°C
LDOPD-7	5’ TTGGCACGGG 3’		70	°C
LDD-10	5’ GGTCTACACC 3’	2988	60	20.5 °C

Table 1.3 Number and size range of bands produced by RAPD-PCR and RAMP-PCR of genomic DNA extracted from four populations of *C. fugax* in Vietnam

Primer code	Total N0 of bands	N0 of polymorphic bands	% of polymorphism
RAPD PRIMERS			
OPA-6	19	19	100.00
OPB-13	10	10	100.00
OPC-2	7	1	14.29
OPC-13	10	8	80.00
OPC-19	9	7	77.78
OPD-2	7	1	14.29
Subtotal 1	**62**	**46**	**74.19**
RAMP PRIMERS			
OPC-2/MS1	9	6	66.67
OPC-2/MS2	15	14	93.33
OPC-2/MS3	12	10	83.33
OPD-2/MS1	8	2	25.00
OPD-2/MS2	10	6	60.00
OPD-2/MS3	12	10	83.33
Subtotal 2	**66**	**48**	**72.73**
Total	**128**	**94**	**73.44**

Table 1.4 Number and size range of bands produced by RAPD-PCR and RAMP-PCR of genomic DNA extracted from four populations of *C. hindsii* in Vietnam

Primer code	Total N0 of bands	N0 of polymorphic bands	% of polymorphism
RAPD PRIMERS			
OPA-2	19	16	84.21
OPA-7	32	30	93.75
OPC-2	17	13	76.47
OPD-2	11	8	72.73
OPD-7	20	14	70.00
OPD-10	21	17	80.95
Subtotal 1	**120**	**98**	**81.67**
RAMP PRIMERS			
OPA-7/MS1	16	10	62.50
OPA-7/MS2	16	11	68.75
OPC-2/MS3	11	9	81.82
OPD-2/MS3	12	9	75.00
OPD-10/MS1	14	8	57.14
Subtotal 2	**69**	**47**	**68.12**
Total	**189**	**145**	**76.72**

Appendix 2 Location and GPS coordinates for sampling sites

Table 2.1 Coordinate location of four *C. fugax* populations

Popul-ation	LV1	LV2	NF	TV
Plant 1	21^0 30’45.6620”N 105^0 13’1.5870”E	210 29’58.1570”N 1050 13’47.9926”E	21^0 31’2.701”N 105^0 13’0.371”E	21^0 26’44.7064”N 105^0 14’57.4322”E
Plant 2	21^0 30’45.6620”N 105^0 13’1.5870”E	21^0 29’56.7121”N 105^0 13’48.6063”E	21^0 31’6.92”N 105^0 13’1.343”E	21^0 26’44.7064”N 105^0 14’57.4322”E
Plant 3	21^0 30’45.6620”N 105^0 13’1.5870”E	21^0 30’3.6008”N 105^0 13’40.8756”E	21^0 31’9.75”N 105^0 12’58.265”E	21^0 26’44.7064”N 105^0 14’57.4322”E
Plant 4	21^0 30’45.6620”N 105^0 13’1.5870”E	21^0 30’8.971”N 105^0 13’41.7”E	21^0 31’7.683”N 105^0 12’53.805”E	

Table 2.2 Coordinate location of four *C. hindsii* populations

Population	PT	HN	HB	HP
Plant 1	21^0 24’30.4318”N 105^0 18’24.4721”E	$21^0$6’24.4650”N $105^0$21’51.6357”E	20°39'53.0"N 105°40'12.1"E	20^0 50’18.953”N 106^0 58’54.25”E
Plant 2	21°20'24.3"N 105°22'48.9"E	$21^0$6’24.4650”N $105^0$21’51.6357”E	20°39'57.0"N 105°40'07.8"E	20^0 47’37.321”N 107^0 3’8.86”E
Plant 3	21^0 24’30.4318”N 105^0 18’24.4721”E	20^0 48’9.3177”N 105^0 38’43.6160”E	20°40'02.3"N 105°40'04.4"E	20°47'56.5"N 106°59'39.3"E
Plant 4	21^0 24’38.6221”N 105^0 18’36.6657”E	ndix	$20^0$53’59.89”N $105^0$23’19.30”E	

Appendix 3 Morphological characteristics and statistics for enzymes and metabolites

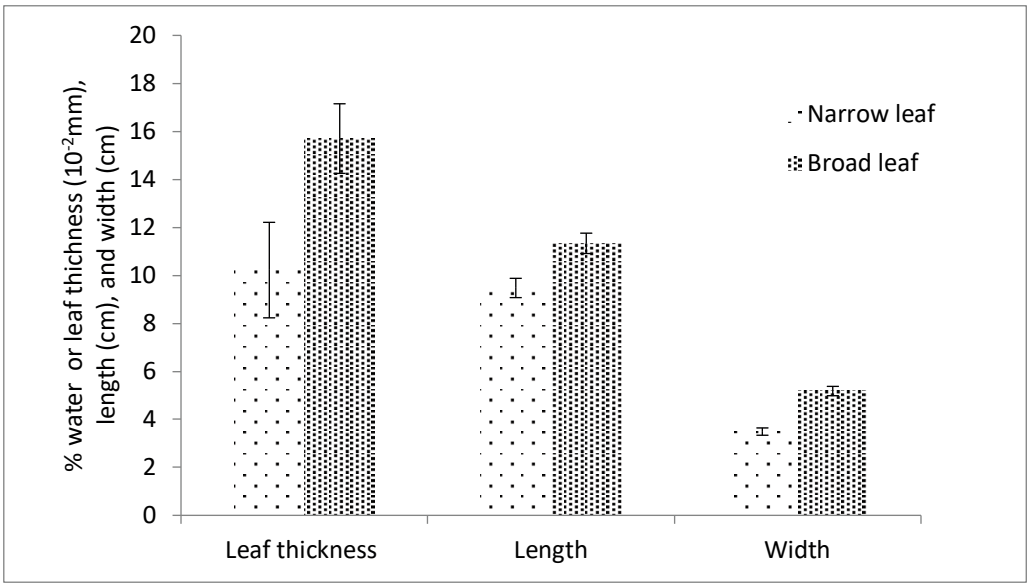


Figure 3.1 Morphological leaf characteristics differences between NL and BL

Table 3.1 Acid and neutral protease activity in NL and BL of *C. hindsii* on ADW (µmol/h/g) basis and protein basis (µmol/h/mg)

Basis	Leaf type	Acid protease			Neutral protease		
		Mean ± SE	F	P	Mean ± SE	F	P
ADW	NL	5.45 ± 0.25	65.14	0.000	1.42 ± 0.11	0.29	0.000
	BL	4.37 ± 0.27			1.35 ± 0.07		
Protein	NL	1.71 ± 0.15	116.12	0.000	0.63 ± 0.08	4.05	0.06
	BL	2.33 ± 0.11			0.62 ± 0.09		

Table 3.2 Proline and glutathione content on ADW (µmol/g and nmol/g, respectively) basis and protein basis (µmol/mg and nmol/mg, respectively) in NL and BL of *C. hindsii*

Basis	Leaf type	Proline			Glutathione		
		Mean ± SE	F	P	Mean ± SE	F	P
ADW	NL	12.71 ± 0.77	3.04	0.103	27.65 ± 2.51	0.86	0.370
	BL	9.7 ± 0.78			29.86 ± 2.72		
Protein	NL	6.26 ± 0.67	235.07	0.000	12.41 ± 1.16	0.91	0.360
	BL	6.23 ± 0.72			12.97 ± 1.18		

Table 3.3 Tocopherol and flavonoids content on ADW basis (µg/g and mg/g, respectively) and protein basis (µg/mg and mg/mg, respectively) in NL and BL of *C. hindsii*

Basis	Leaf type	Tocopherol			Flavonoid		
		Mean ± SE	F	P	Mean ± SE	F	P
ADW	NL	18.81 ± 1.5	9.34	0.008	18.75 ± 0.96	128.76	0.000
	BL	22.29 ± 1.23			23.16 ± 1.03		
Protein	NL	6.77 ± 1.2	2.34	0.149	7.43 ± 0.52	69.56	0.000
	BL	9.79 ± 1.1			8.78 ± 0.52		

Table 3.4 Phospholipids and MDA content on ADW basis (µg/g and nmol/g, respectively) and protein basis (µg/mg and nmol/mg, respectively) in NL and BL of *C. hindsii*

Basis	Leaf type	Phospholipids			MDA		
		Mean ± SE	F	P	Mean ± SE	F	P
ADW	NL	6.25 ± 0.7	89.36	0.000	4.7 ± 0.89	33.61	0.000
	BL	8.23 ± 0.69			2.53 ± 0.4		
Protein	NL	3.83 ± 0.56	0.009	0.923	2.3 ± 0.53	39.88	0.000
	BL	4.2 ± 0.67			1.09 ± 0.23		

Appendix 4 Raw computer data for proteomics

Table 4.1 Differentially expressed proteins with their significant biological process and molecular function (proteins in the Italic present in both NL and BL with the same density, while other proteins were identified upregulated (> 1.5 folds, $p < 0.05$) and only present in BL)

Protein in the biological process	Molecular function
Signalling transduction and defence response	
Germin-like protein subfamily 3 member 1	Manganese ion binding, nutrient reservoir activity
Kunitz family trypsin ad protease inhibitor 2-like	Protein phosphatase inhibitor activity; ATPase inhibitor activity; endopeptidase inhibitor activity
Leucine-rich repeat protein 1 (somatic embryogenesis receptor kinase 1 isoform X1)	Transmembrane receptor protein serine/threonine kinase activity (EC:2.7.11)
Auxin-binding protein ABP19a	Manganese ion binding; nutrient reservoir activity
Probable serine/threonine-protein kinase PIX13 (probable disease resistance)	Transmembrane receptor protein serine/threonine kinase activity (EC:2.7.11).
Protein IQ-DOMAIN 1 isoform X1	Calmodulin binding, microtubule binding
Disease resistance protein (CC-NBS-LRR class) family	ADP binding, protein binding
Acidic endochitinase	Chitinase activity (EC:3.2.1.14); chitin binding
Protein modification and catabolic process	
S locus-linked F box protein	Protein binding in protein S-linked glycosylation via cysteine.
Leucine-rich repeat protein 1 (somatic embryogenesis receptor kinase 1 isoform X1)	Protein binding
F-box protein At1g70590	Protein binding
Beta-1,3-galactosyltransferase 14	Galactosyltransferase activity
Cullin-4	Ubiquitin protein ligase binding
Probable serine/threonie-protein kinase PIX13	Protein binding, ATP binding, ADP binding,
F-box/FBD/LRR-repeat protein	Protein binding
Autophagy-related protein 8D-like	Protein binding
Elongation factor G family protein	Translation elongation factor activity, GTPase activity (EC:3.6.1.15), GTP binding
Nucleic acid metabolic process	
Staphylococcal Nuclease domain-containing protein 1	Nuclease activity

Protein in the biological process	Molecular function
Retrovirus-related Pol polyprotein from transposon 297 family	Nucleic acid binding; RNA-DNA hybrid ribonuclease (EC:1.4.3.1); D-aspartate oxidase activity, oxidoreductase activity
Pentatricopeptide repeat-containing protein At4g21065	RNA binding
Cullin-4	Ubiquitin protein ligase binding
Endoribonuclease Dicer homolog 3a	Ribonuclease III activity, protein binding, ATP binding, DNA binding, hydrolase activity; endoribonuclease activity, producing 5'-phosphomonoesters (EC:3.1.30; EC:3.1.26; EC:3.1.26.3)
Putative nuclease HARBI1	Nuclease activity
Aromatic compound biosynthetic process	
Leucoanthocyanidin reductase 1	Leucoanthocyanidin reductase (EC:1.17.1.3)
F-box/kelch-repeat protein At5g51250-like	Ubiquitin-protein ligase binding
Energy and photosynthesis, and protein transport	
ATP synthase subunit D, mitochondrial	Proton-transporting ATP synthase activity (EC:3.6.1.3; EC:3.6.1.15)
ATP synthase CF1 alpha subunit (chloroplast) ADD31493.1	ATP biding (EC:3.6.1.3; EC:3.6.1.15)
Chlorophyll a-b binding protein of LHCII type 1	Chlorophyll binding, pigment binding, metal ion binding
Ycf1	Protein transporter activity
Cystinosin homolog	L-cystine transmembrane transporter activity
Localization, Inorganic cation transmembrane transport	
ATP synthase CF1 alpha subunit (chloroplast) ADD31493.1	Proton-transporting ATP synthase activity, rotational mechanism (EC:3.6.1.3; EC:3.6.1.15)
Hypothetical protein AXG93_1154s1520	Voltage-gated potassium channel activity
Inorganic phosphate transporter 1-4 like	Inorganic phosphate transmembrane transporter activity
Transportin MOS14	Nuclear localisation sequence binding, protein transporter activity
Protein trichrome birefringence-like 10	O-acetyltransferase activity
ATP synthase subunit D, mitochondrial	Proton-transporting ATP synthase activity, rotational mechanism (EC:3.6.1.3; EC:3.6.1.15)
Others	

Protein in the biological process	Molecular function
Myosin-2-like isoform X1	Actin-binding, ATP binding, motor activity (EC:3.6.1.15)
Uncharacterised protein LOC105059940 isoform X2	
Hypothetical protein VITISV_000703	Protein malonylation
Hypothetical protein VOLCADRAFT	

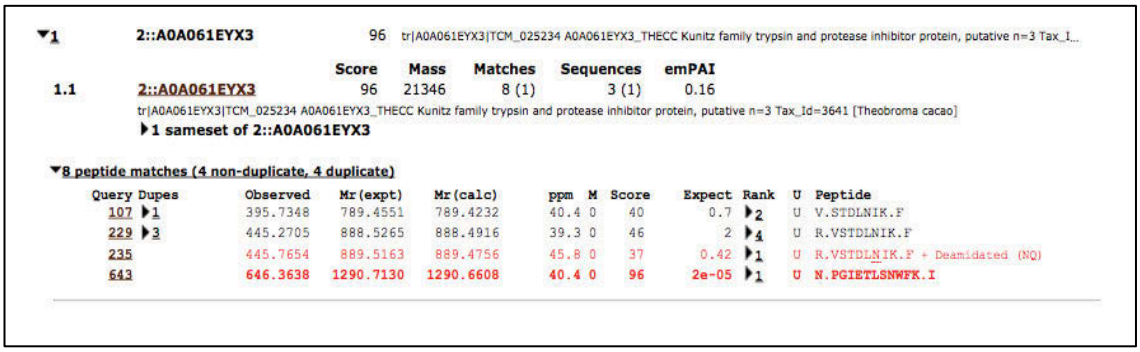


Figure 4.1 Identification of proteins by LC-MS/MS analysis using MASCOT tool

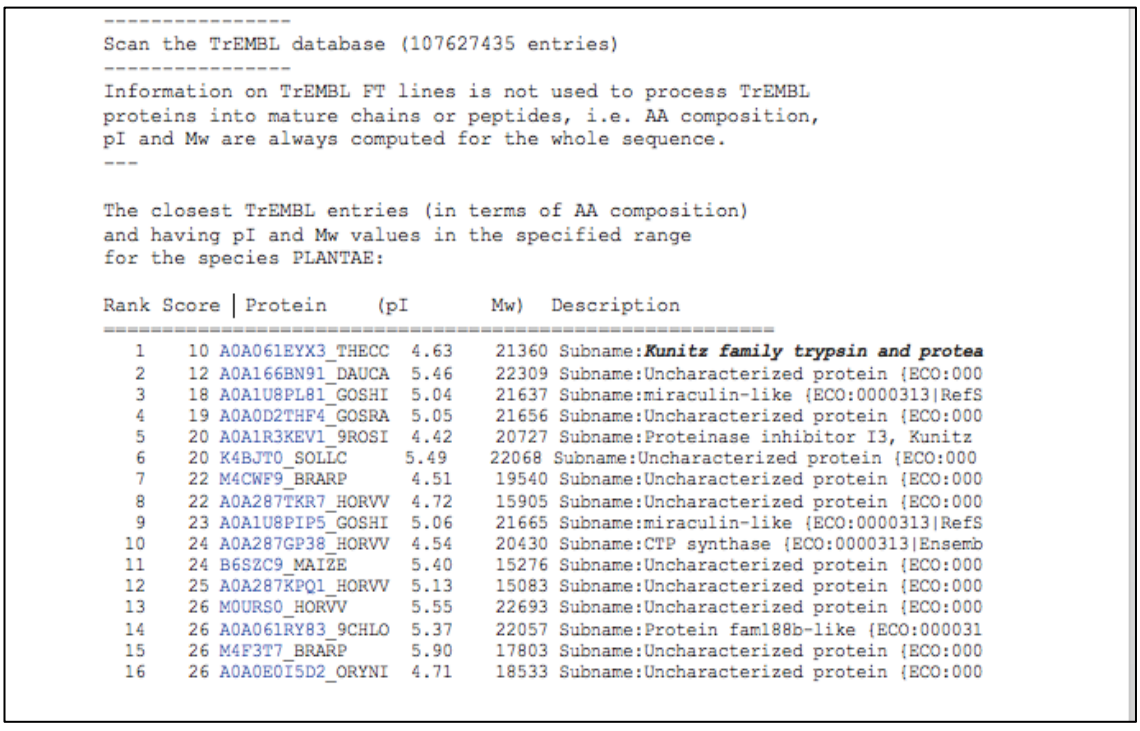
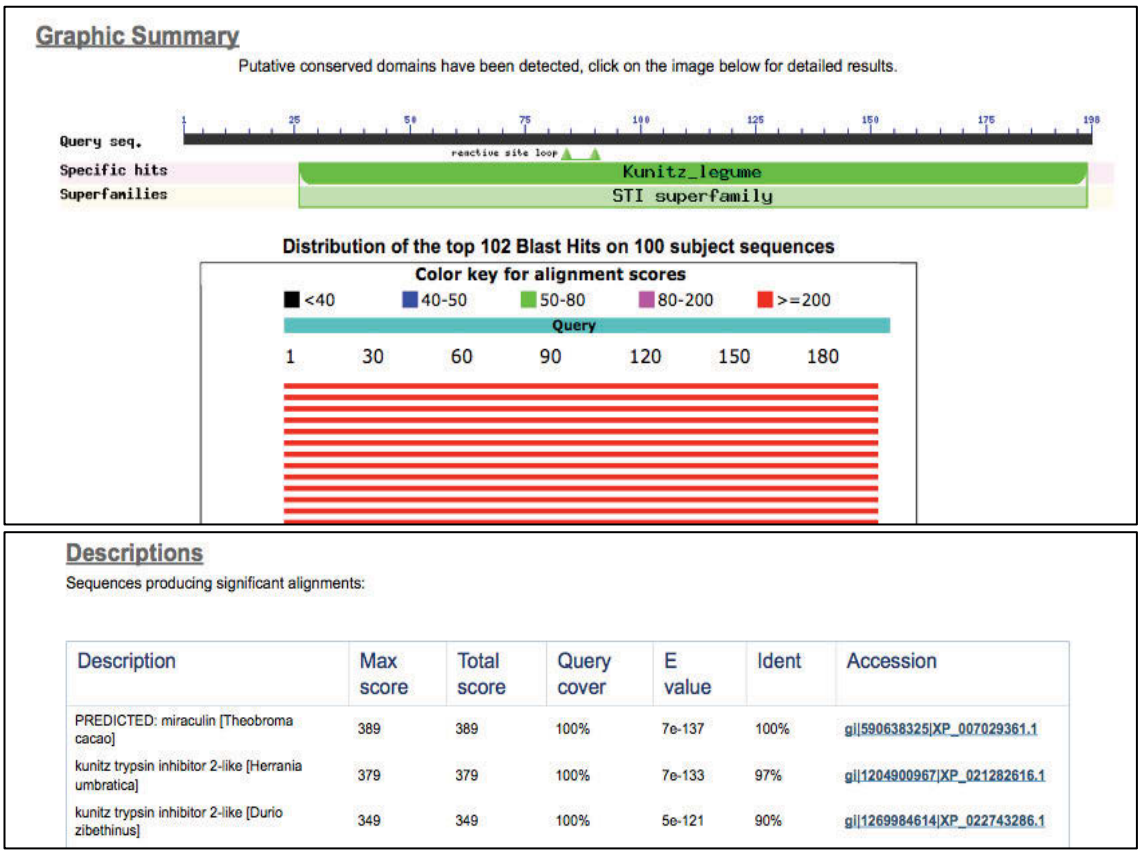


Figure 4.2 Identification of proteins by peptide fingerprinting data using MultiIdent tool



Descriptions

Sequences producing significant alignments:

Description	Max score	Total score	Query cover	E value	Ident	Accession
PREDICTED: miraculin [Theobroma cacao]	389	389	100%	7e-137	100%	gi|590638325|XP_007029361.1
kunitz trypsin inhibitor 2-like [Herrania umbratica]	379	379	100%	7e-133	97%	gi|1204900967|XP_021282616.1
kunitz trypsin inhibitor 2-like [Durio zibethinus]	349	349	100%	5e-121	90%	gi|1269984614|XP_022743286.1

Figure 4.3 Putative conserved domain of Kunitz protein and top hit with e-value detected by using BLAST for comparing the similarity with the protein database of NCBI