

# Comparative proteomic analysis of *Celastrus hindsii* Benth. phenotypes reveals an intraspecific variation

Van Huy Nguyen · Thanh Loan Pham · Thi Tam Tien Ha · Thi Le Thu Hoang

Received: 4 November 2020 / Revised: 2 December 2020 / Accepted: 2 December 2020  
© Korean Society for Plant Biotechnology

**Abstract** In Vietnam, *Celastrus hindsii* Benth, a medicinal plant rich in secondary metabolites, has been used to alleviate distress caused by ulcers, tumors, and inflammation for generations. The occurrence of two phenotypes, Broad Leaf (BL) and Narrow Leaf (NL), has raised questions about the selection of appropriate varieties for conservation and crop improvement to enhance medicinal properties. This study examined molecular differences in *C. hindsii* by comparing protein profiles between the NL and BL types using 2D-PAGE and MS. Peptide sequences and proteins were identified by matching MS data against the MSPnr100 databases and verified using the MultiIdent tool on ExPASy and the Blast2GO software. Our results revealed notable variations in protein abundance between the NL and BL proteomes. Selected proteins were confidently identified from 12 protein spots, thereby highlighting the molecular variation between NL and BL proteomes. Upregulated proteins in BL were found to be associated with flavonoid and amino acid biosynthesis as well as nuclease metabolism, which probably attributed to the intraspecific variations. Several bioactive proteins identified in this study can have applications in cancer therapeutics. Therefore, the BL phenotype characterized by healthier external morphological features has higher levels of bioactive compounds and could be better suited for medicinal use.

**Keywords** Bioinformatics, *Celastrus hindsii*, Comparative analysis, Medicinal plant, Morphological variation, Mass spectrometry, proteomics

## Introduction

The medicinal plant *Celastrus hindsii* Benth. has been utilized as a traditional medicinal source for treatment of some chronic disorders in Northern Vietnam, particularly by Muong people. Over past decades, *C. hindsii* has been discovered for a variety of chemical components, including sesquiterpenes, triterpenes, alkaloids, and flavonoids, as well as their bioactivities (Hu et al. 2014; Thuy et al. 2007; Viet et al. 2019). Other antioxidants (vitamin C, vitamin E, and some phenolics), which have been shown to have some health effects on reducing the occurrence of cancer and cardiovascular disorders, were also isolated from dried leaves of *C. hindsii* (Ly et al. 2006). The rapid increase in the global population with chronic diseases leads to an elevated uptake of these medicinal compounds. With a relatively high content of diversely specialized secondary metabolites, *C. hindsii* is attracting greater international scholarly attention. Recent phytochemical investigations conducted on *C. hindsii* have based on plant materials collected from different locations. Therefore, the results have minimal comparative implications for the selection of appropriate varieties for conservation and medicinal plant crop improvement. The occurrence of two different leaf morphological phenotypes (narrow-leaf - NL and broadleaf - BL) grown in both wild and cultivated conditions may be the result of diverse factors, including molecular regulation.

Plants respond to changing environmental conditions by reprogramming their cellular machinery at the molecular levels (Amudha et al. 2005). In other words, the biosynthesis of plant secondary metabolites is induced by external factors and regulated by complicated molecular mechanisms. Profiling a proteome of an organism can provide insights into the cellular pathways of that species, particularly the biosynthetic pathway of specialized secondary metabolites. In the light of recent plant genomic studies, the power of proteomics by differential displays of

V. H. Nguyen (✉) · T. L. Pham · T. T. T. Ha · T. L. T. Hoang  
Institute of Applied Research and Development, Hung Vuong  
University, Phu Tho, Vietnam  
e-mail: huytnhv@gmail.com

the proteome has allowed the identification of polypeptides by a mass spectrometer (MS) (Anguraj 2015). Up to date, there has not been any proteome report from medicinal plant *C. hindsii*. Therefore, the underlying cellular mechanisms leading to the divergence 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 examine differentially expressed proteins under various conditions and in different plant cultivars and species (De Filippis and Magel 2012; Wang et al. 2015). However, the proteome profiling of the variation in *C. hindsii* phenotypes has not yet been investigated. Findings from previous research could be useful to further understand the pharmaceutical properties of *C. hindsii* - a potential candidate for medicinal plant crops.

Therefore, the current study, for the very first time to the best of our knowledge, investigated further molecular changes in the plant *C. hindsii* to disclose the protein profile in NL and BL. A comparative analysis using 2D-PAGE, coupled with MS, was adopted to discover the differential proteins corresponding to the morphological formations of NL and BL in *C. hindsii*. Combining proteomics with molecular approaches allows a better understanding of the biosynthesis pathway of valuable secondary metabolites of medicinal values.

## Materials and Methods

### Study area

This research component was a part of the 4-year project in which proteomics component was conducted from 12/2017 to 06/2018. in the biological laboratory of the Life Sciences Department, Faculty of Science, University of Technology Sydney, Australia.

### Plant materials

Plant materials were collected in home gardens in Phu Tho Province, Vietnam. 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 were then store in  $-80^{\circ}\text{C}$  freezer. Air-dried leaf was weighed using an electronic balance (Sartorius, Quintix224-1S, Germany, accuracy  $\pm 0.0001$  g), sliced into thin pieces ( $1\sim 2$  mm<sup>2</sup>), and frozen immediately in liquid nitrogen. The frozen tissues were subsequently ground into fine powder by pul-

verizing using a cryomill (Retsch MM200, Rheinische, Germany) with a 1 cm stainless steel ball. The final powder was then stored in liquid nitrogen for immediate protein extraction or stored at  $-80^{\circ}\text{C}$  for later extraction.

### Chemicals

Acrylamide, polyvinylpyrrolidone (PVPP), standard protein molecular weights, and carrier ampholytes were purchased from Sigma (St. Louis, MO, USA). SDS, TEMED, ammonium acetate, and  $\beta$ -mercaptoethanol were from Sigma-Aldrich. The immobilized pH gradient, chemical agents, and equipment required for 2-DE were provided by Bio-Rad. All solutions were prepared using double-distilled water.

### Protein extraction

The borax-PVPP-Phenol method (BPP) was proven to be effective in dealing with recalcitrant plant tissues and facilitate downstream applications by removing most of the interfering compounds and producing high-quality protein samples (Wang et al. 2007). Interfering compounds such as polysaccharides, polyquinones, and phenolic compounds were removed by borax and PVPP.  $\beta$ -mercaptoethanol and ascorbic acid was 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 sulfate saturated-methanol) were replaced by 0.1 M ammonium acetate saturated-methanol. These alterations facilitated protein solubility and ensured a minimum activity of protein degradation.

In short, 1g of frozen dry leave powder was 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%  $\beta$ -mercaptoethanol v/v and 30% sucrose w/v. After 5 min vortexing at room temperature, two volumes of Tris-saturated phenol (pH 8.0) were added, and then vortexed for further 10 min. After being centrifuged ( $4^{\circ}\text{C}$ , 15 min,  $15,000 \times g$ ), the upper phase was transferred to a new centrifuge tube then added an equal volume of extraction buffer. After vortexing the mixture for 10 min and centrifugation (at the same condition), the supernatant was then collected. The precipitation of proteins was carried out by adding five folds of 0.1 M ammonium acetate saturated-methanol and incubated at  $-20^{\circ}\text{C}$  overnight or at least six h. The protein pellet collected after centrifugation 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°C 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 disulfide bonds in a single step, for 90 min at room temperature, using the reducing agent tributylphosphine (TBP, 5 mM) and an alkylating acrylamide monomer (AM, 20 mM). The reaction was quenched using dithiothreitol (DTT, 20 mM).

### Quantification of proteins

Protein concentration was determined using the 1D PAGE and densitometry. 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.

### Protein fractionation and detection

2D-PAGE was carried out following De Filippis & Magel (De Filippis and Magel 2012) with some modifications. Proteins (300 µg) were analyzed using IEF in the first dimension, which was based substantially on their charge and pH equilibrium. Immobilized pH gradient (IPG) strips (Bio-Rad, pH 3-10, 11 cm) were passively rehydrated using rehydration solution (UTC7) at room temperature for 6 h at least. 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°C with a Protean IEF device (Bio-Rad). The gel strips were subsequently equilibrated in equilibration buffer (6 M urea, 250 mM 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™ precast polyacrylamide gel (Bio-Rad) in MES SDS running buffer (Invitrogen, Life Technologies) in a Midi format electrophoresis systems (Bio-Rad Laboraroties, NSW, Australia) using the 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 (Bio-Rad Laboraroties, NSW, Australia). Gels were then de-stained, an image obtained using a fluorescence scanner (Typhoon FLA-3500, Freiburg, Germany).

The gel was calibrated in the vertical sodium dodecyl sulfate (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 were examined had to be present in a minimum of three separate 2D gels out of five runs. A total of six 2-DE gels obtaining from three individual replicates of each phenotype group were examined. Up- and down-regulated proteins referred to differentially regulated proteins in NL samples compared to BL samples.

### Protein identification

The excised spots from the 2D gels were digested with trypsin (Trypsin gold-MS grade Promega, Mannheim, Germany), followed by LC/MS/MS analysis by the guidelines of Pokharel et al. (2016). Briefly, the gel was sectioned into 1 × 1 mm pieces, then de-stained by 50% acetonitrile (ACN)/50 mM NH<sub>4</sub>HCO<sub>3</sub> in incubated condition for 10 min at ambient temperature. When the stain disappeared, 100% ACN was added to dehydrate the gel pieces. Rehydration was carried out with 12.5 ng/µl of trypsin solution and incubated at 37°C overnight. The supernatant was collected and sonicated for 10 min in a water bath. The solution was added with 30 µl of 50% CAN/2% formic acid, and the volume was reduced to 15 µl by rotary evaporation (IKA, Staufen, Germany). The peptide solution was centrifuged (14,000 x g for 10 min) to eliminate any insoluble material before LC/MS/MS analysis, as described by (Kumar et al. 2017).

### Bioinformatics

Peptides and proteins were identified using both MS and PEAKS Studio software (Peak Studio 7.5, Bioinformatics Solution Inc., Waterloo, ON, Canada). The MS and MS/MS data generated by the QSTAR were searched restrictedly within *Viridiplantae* (green plants) by the Mascot Daemon (version 2.4) and the MSPnr100 databases containing 75,925,788 sequences; 27,045,014,025 residues). The results were filtered by counting only protein hits with at least one unique peptide (Bold Red) and excluding peptide hits with a p > 0.05. A minimum Mascot score of 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<sup>14-16</sup>. The threshold was set to a minimal significance of 1e-3 and an identity

**Table 1** Protein yield from different plant species extracted using BPP

Plant species	Tissue	Protein yield ( $\mu\text{g/g}$ FW)	Note
<i>Celastrus hindsii</i>	Leaf, fully expanded	2347 $\pm$ 484	Medicinal plant
<i>Arabidopsis thaliana</i>	Leaf, fully expanded	2250 $\pm$ 321	Model plant
<i>Prunus persica</i>	Leaf, young	2176 $\pm$ 288	Woody plant
<i>Pinus bungeana</i>	Needle, mature	1960 $\pm$ 352	Recalcitrant plant

percentage of >75%. Further protein identification was performed, employing mass fingerprint data, pI, and molecular weight with the MultiIdent tool in ExPASy ([www.expasy.org/tools/](http://www.expasy.org/tools/)) (Wilkins and Williams 1997; De Filippis and Magel 2012) to search for close matching protein across species. The theoretical pI and molecular weight of the BLAST hit were calculated using the ExPASy tool ([http://web.expasy.org/compute\\_pi/](http://web.expasy.org/compute_pi/)).

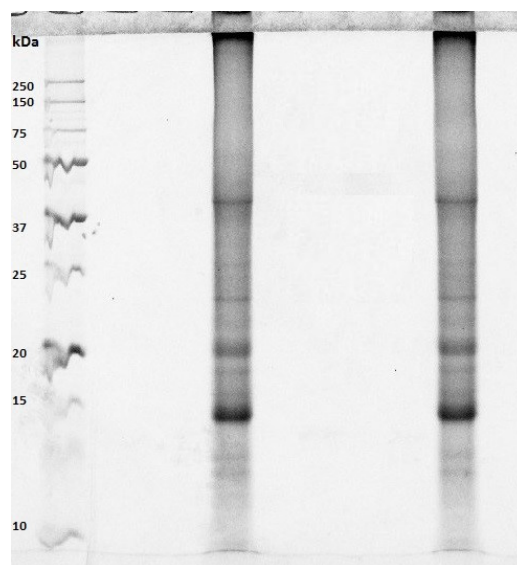
Due to experimental variations caused by instrumental uncertainty in operation, as often experienced, not all spots were detected on each gel of the same sample. Five gels were made for each sample to average out this experimental variation. Only spots presented in at least three gels were chosen. All selected spots were successfully characterized by LC-MS/MS and identified by the Mascot to search proteins against the MSPnr100 databases. Due to the extremely poor genome and the availability of protein sequence information on *C. hindsii* and the family Celastraceae, the identification relied almost exclusively on homology to proteins identified from other well-characterized plant species. Therefore, a comprehensive strategy for database searching was developed with three conditions. If a significant score was obtained for one or more peptides, an error-tolerant search was performed on the sequence of the identified protein to look for additional homologous peptides containing a modification, e.g., an amino acid substitution.

The determination of the identified proteins' biological role and function was carried out by assigning them to Gene Ontology (GO) using Blast2GO software, coupled with the UniProt GO annotation program. The GO database, BLAST annotations, and information published in the literature were also used to analyze and classify the identified proteins based on their cellular localization, molecular functions, and biological processes. Based on the searches, specific up-regulated and down-regulated proteins were determined.

## Results and Discussion

### Protein extraction

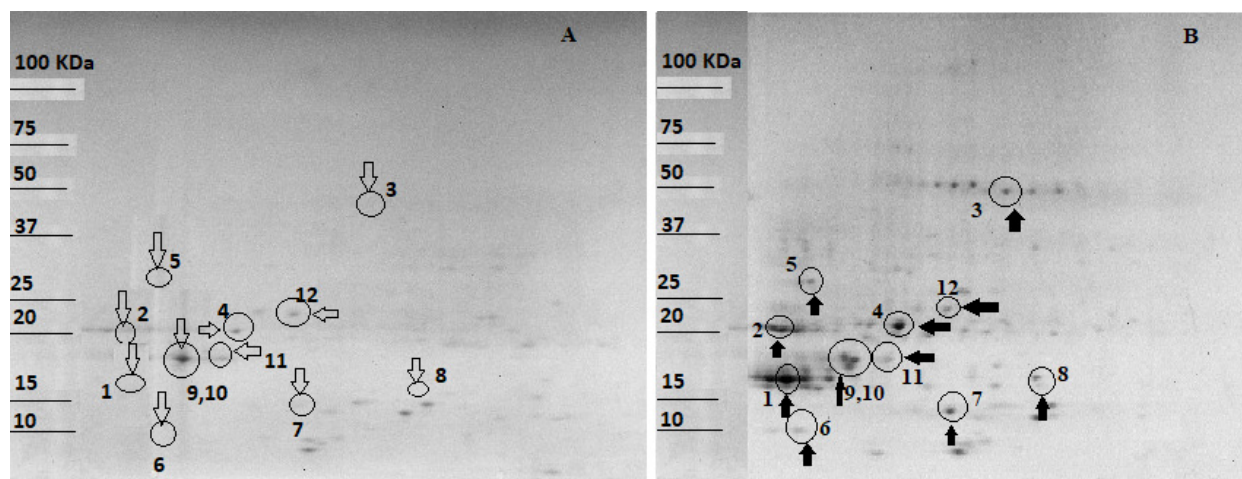
Protein yields from BPP of leaf tissues of four species,



**Fig. 1** Protein profile: 1-D gel with two replicates showed the abundance of low molecular mass proteins but lack of high molecular ones

including *C. hindsii* extracted from this experimental study and another three plant species by Wang et al. (2013), are listed in Table 1. The average protein yield in *C. hindsii* was 2347  $\pm$  484 mg/g fresh weight, which is consistent with the previous report by Wang et al. (2013), although it was a slightly higher than the other plants. This higher yield is probably due to the modification, including the addition of SDS and protease inhibitors. Borax/PVPP/Phe extraction, which integrated the advantages of phenol-based protocols and TCA/Acetone methods, has shown to be the method of choice as it was competent to deal with recalcitrant leaf tissues (Li et al. 2015; Wang et al. 2013). Besides, the other plant species, particularly woody and recalcitrant plants, may contain more substantial amounts of polysaccharides, polyphenols, which may obstruct protein isolation and produce less yield (Alam et al. 2013).

Using the BPP method, proteins extracted from *C. hindsii* leaves were distinct between medium and lower molecular weight ( $M$ ) regions of the gel (Fig. 1). However, bands at a 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 smaller



**Fig. 2** Protein distribution and profiling of *C. hindsii* expressed in NL (A) and BL (B) phenotypes. (up-pointing arrow: upregulated; leveled arrow: present in both NL and BL; down-pointing arrow: downregulated)

proteins were enriched (Carpentier et al. 2005).

#### Proteomic profiling

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

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. The analysis of the proteome visualized in 2-DE gels (pI 4–7) discovered a total of 8 spots, which changed in volume variation (1.5-fold,  $p < 0.05$ ) in BL compared to the NL. Two spots newly appeared in BL, and the other two significant spots, which were expressed with high density in the 2D gels of both NL and BL, were also selected for identification. By using the database search with the method above, 12 protein spots could have possible identifications (Table 2). By initial searching 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 insufficient 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. As proteins of interest in *C. hindsii* were likely not to be present in the database,

protein identification by peptide mass fingerprinting often requires highly homologous proteins from other species whose proteomes are available in the databases.

By using the database search with the method above, 12 protein spots could have possible identification (Table 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 the 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 often 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 could not be adequate to ensure the identification because the conservation of homology is often not guaranteed at the tryptic digest level.

Despite this, the MS/MS fingerprinting method used in this study was successful in several cases. The use of mass fingerprinting peptide data using the MultiIdent tool in ExPASy at ([www.expasy.org/tools/](http://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

**Table 2** Proteins classified as high and low score based on Mascot and BLAST search scores. Proteins with the highest scores were categorized as high score; the second highest scores with significant e-value and identity were categorized as low score. \*\* rating: High scored pharmaceutical proteins

Spot no.	Mascot score	Protein name	Regulation	BLAST Score	E-value	Identity*
High score						
1	40	Inorganic phosphate transporter	UR	45.2	1e-05	93%
2	99	Acidic endochitinase	UR	560	0.0	100%
3	823	ATP synthase CF1 alpha subunit	UR	827	0.0	100%
4	86	Chlorophyll a-b binding protein	PB	839	0.0	100%
5	29	F-box/FBD/LRR-repeat protein	UR	53.2	3e-07	100%
6	24	Ycf1 (chloroplast)	UR	91.8	5e-20	100%
7	20	Myosin-2-like isoform X1	UR	60	2e-09	100%
8	26	F-box protein At1g70590	UR	53.7	3e-07	100%
9	96	Kunitz trypsin protease inhibitor **	UR	389	7e-137	100%
	55	Cullin-4	UR	1509	0.0	100%
10	66	Pentatricopeptide repeat-containing	UR	985	0.0	100%
	49	Disease resistance protein RPS4-like	UR	1488	0.0	100%
11	108	Germin-like protein 11	PB	360	9e-125	100%
	54	ATP synthase subunit d	PB	985	0.0	100%
12	33	Protein trichome birefringence-like	PB	1483	0.0	100%
Low score						
1	31	F-box/kelch-repeat protein	UR	40.9	9e-08	100%
	35	IQ-DOMAIN 1-like	UR	40.5	3e-04	100%
5	15	Beta-1,3-galactosyltransferase 14	UP	54.5	2e-07	100%
6	20	Retrovirus-related Pol polyproteins from transposon 297 family	UR	57.9	1e-08	100%
	24	Endoribonuclease Dicer homolog **	UR	49.4	5e-06	100%
8	22	Staphylococcal Nuclease domain-containing protein 1 **	UR	50.3	4e-06	100%
	64	S locus-linked F box protein	UR	723	0.0	100%
9	53	Leucoanthocyanidin reductase 1	UR	662	0.0	100%
	60	Germin-like protein 11	UR	360	9e-125	100%
10	33	Elongation factor G family protein	UR	1302	0.0	100%
	54	Auxin-binding protein ABP19a	PB	341	1e-118	100%
11	51	Autophagy-related protein 8D-like	PB	412	2e-145	100%

\* The cut-off for query cover is 82% and identity is 75%, however, most of proteins have 100% of these two values. UR: Upregulated; PB: Present in both.

whose scores are less than 30, and their ranks are far different from the sequential matchings.

The identified peptides by Mascot were additionally confirmed 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. Besides, 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 one was only found in protein nine and elongation factor G family protein was only identified in protein 10.

#### Proteins involved in the amino acid turnover

In *C. hindsii*, S locus-linked F-box protein, F-box/FBD/LRR-repeat protein, and Cullin-4 were found up-regulated in BL. These three proteins are all important components of cullin-ring ubiquitin ligase systems. Recent studies have revealed that Cullin-4 can target histones for ubiquitination. The histone ubiquitination may facilitate the cellular response to DNA damage (Dai and Wang 2006). Guo et al. (2013) reported that Cullin-4 played a key role in DNA repair and damages due to genotoxic stress. Roodbarkelari et al. (2010)



confirmed that Cullin-4 acted as a major regulator of endoreplication in *Arabidopsis* trichomes. While endoreplication has been found to play a developmental role in controlling cell size, regulating cell morphology, and increasing DNA content is typically associated with larger cells (Harashima and Schnittger 2010). As a result, an increased level of cullin-4 probably explains its role in the formation of the broader and thicker leaf blade in some variants of *C. hindsii*.

Kelch containing F-box (KFB) proteins were expressed up-regulated in BL (protein 1), possibly indicating their roles in differentiating NL and BL. In *Arabidopsis*, overexpression of CFK1 (a plant-specific Kelch repeat-containing F-box protein) produced phenotypes with elongated hypocotyls, DNA content, and increased ploidy (Franciosi et al. 2013; Miyazaki et al. 2011). Chen et al. (2013) reported that KFB from rice plays a role in the regulation of leaf senescence, grain number, and seed size. The disturbance of cytokinin translocation, distribution, or signal perception in plants severely degrades their vascular formation and lignification (Li et al. 2011; 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*.

#### Proteins interfering nucleic acid metabolism

##### *Staphylococcal nuclease (TSN) domain-containing protein (up-regulated in spot 8)*

In plants, the molecular function of TSN remains poorly characterized, mostly as described in *Arabidopsis* (Vanderschuren et al. 2013). TSN is conserved in all studied 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. hindsii* may indicate its role in promoting stress tolerance or protect the tissues from cell death. 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 cancer (Xing et al. 2018).

##### *Endoribonuclease Dicer homolog 3a (up-regulated in spot 8)*

Dicer, one of the core proteins within the RNAi was found upregulated in BL. It is highly conserved from archaea to eukaryote and expands largely in plants and animals with multiple members of the family (Jia et al. 2017). Dicer is an important endonuclease in the biogenesis of miRNAs. The up-regulation of Dicer in *C. hindsii* leaf extract was probably associated with the biosynthetic activity of sRNAs. Other studies also revealed that miRNAs involving in different metabolic processes mainly played a role in the regulation of essential oil pathways in *Mentha spp*, biosynthesis of gingerol in *Zingiber officinale* (Singh et al. 2016), and also regulated terpenoid backbone, isoquinoline alkaloids, and flavonoids biosynthesis in *Curcuma longa* (Singh and Sharma 2017). The study by Witwer & Hirschi (2014) has 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. 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.

##### *Retrovirus-related Pol polyprotein from transposon 297 family (up-regulated in spot 6)*

The results of the present study revealed that retrovirus-related pol polyprotein from transposon 297 family (encode retrotransposon) were differentially expressed in BL of *C. hindsii*. In other genomic and proteomic studies on plants, transposable elements were differentially expressed in response to a diverse array of activating stimuli (Galindo-González 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 up-regulated in *Quercus suber* by wounding (Rocheta et al. 2012), elevated by fungal elicitors, wounding, viral inoculation in *Nicotiana tabacum* (Anca et al. 2014), up-regulated in response to heat stress, aphid infection, and salicylic acid (Voronova et al. 2014). The presence and upregulation of retrovirus-related pol polyprotein might suggest its importance in the defense system of *C. hindsii* in response to biotic and abiotic stresses.

##### *Pentatricopeptide repeats (PPR)-containing protein*

Pentatricopeptide repeats (PPR) were found to be present in three spots and strongly up-regulated in BL of *C. hindsii*.

It might indicate their vital role in the healthier growth of BL phenotype and probably related to the accumulation of pigments as dark purple leaf color was only found in BL. PPRs are of the most popular repeat domain proteins, which are reported to play several functional roles in growth, development, and stress response in plants (Mandal et al. 2018). 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 were found to be associated with leaf functioning and development, and tolerance to stress conditions (Barkan and Small 2014; Zhang et al. 2017). Furthermore, PPR40 can lessen the generation of reactive oxygen species (ROS) by protecting plants via reducing oxidative damage during stress (Zsigmond et al. 2008). The purple leaf color in the medicinal plant is often associated with the accumulation of anthocyanins in plants, which also show to have antibacterial activity against some pathogens (Skaar et al. 2014). Therefore, the up-regulated level of PPRs in BL might facilitate this phenotype in growth and pigment accumulation, probably anthocyanins.

#### Secondary metabolite biosynthesis

##### *Leucoanthocyanidin reductase 1 (up-regulated in spot 9)*

Proanthocyanidins (PAs, also known as condensed tannins) are ubiquitous in the species in Caprifoliaceae, Celastraceae, and Theaceae (Jiang et al. 2017). In PAs biosynthetic pathway, leucoanthocyanidin reductase (LAR) has been shown to convert leucocyanidin to (+)-catechin and was observed up-regulated in BL. It is commonly agreed in recent studies that LAR is a positive regulator of PAs biosynthesis. Wen et al. (2015) examined the effect of postharvest UV irradiation on flavanol polyphenol accumulation in the grape berry. It indicated that the flavanol polyphenol reached its highest value when exposed to UV-C irradiation and induced the transcription of the LAR gene and its gene products. BL which was collected for the current proteomic experiments mostly came from organized 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 fertilizing provisions probably restricted the growth of the plant, particularly leaf expansion, and the formation of specialized metabolites such as flavonoids.

Proteins with antitumor / anti-inflammatory activity and therapeutic applications

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 (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* (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.

Due to the extremely poor genome and the availability of protein sequence information on *C. hindsii* and the family Celastraceae, the identification of all proteins were not satisfied. Nevertheless, the significant molecular variation revealed by proteomic results between NL and BL may be considered as important biomarkers for breeding selection and biotechnological development to define targets for medicinal plant cultivation. However, in further proteomics research, more sample replication over more plant genotypes is required to validate these conclusions.

#### Conclusion

To our knowledge, this work represents the first comparative proteomic investigation *C. hindsii* between NL and BL phenotypes. The changes in protein abundance revealed significant variations between NL and BL proteomes. Some proteins were discussed to highlight the metabolic changes between NL and BL proteomes, particularly associated with amino acid turnover, nuclease metabolism, flavonoids biosynthesis, and medicinal properties. This study may help the researchers to uncover the critical areas of molecular biology that many researchers were not able to carry out. The results also give suggestions for agricultural managers and farmers in cultivating medicinal plant crops, which



can produce higher agronomic and pharmaceutical properties. It can be concluded that the BL phenotype, which was characterized by healthier external morphological features and possibly better medicinal properties, could be chosen for crop development.

## Acknowledgments

The authors are grateful to the Ministry of Education and Training for financial assistance and colleagues from the Institute of Applied Research and Development, Hung Vuong University, Vietnam for technical support during the project.

## References

- Alam I, Sharmin S, Kim K-H, Kim Y-G, Lee J, Lee B-H (2013) An improved plant leaf protein extraction method for high resolution two-dimensional polyacrylamide gel electrophoresis and comparative proteomics. *Biotech Histochem* 88:61-75
- Amudha P, Jayakumar M, Kulandaivelu G (2005) Impacts of ambient solar uv (280-400 nm) radiation on three tropical legumes. *J Plant Biol* 48:284-291
- Anca IA, Fromentin J, Bui QT, Mhiri C, Grandbastien MA, Simon-Plas F (2014) Different tobacco retrotransposons are specifically modulated by the elicitor cryptogein and reactive oxygen species. *J Plant Physiol* 171:1533-1540
- Anguraj Vadivel AK (2015) Gel-based proteomics in plants: time to move on from the tradition. *Front Plant Sci* 6:2-5
- Barkan A, Small I (2014) Pentatricopeptide repeat proteins in plants. *Annu Rev Plant Biol* 65:415-442
- Carpentier SC, Witters E, Laukens K, Deckers P, Swennen R, Panis B (2005) Preparation of protein extracts from recalcitrant plant tissues: An evaluation of different methods for two-dimensional gel electrophoresis analysis. *Proteomics* 5:2497-2507
- Champagne A, Rischer H, Oksman-Caldentey KM, Boutry M (2012) In-depth proteome mining of cultured *Catharanthus roseus* cells identifies candidate proteins involved in the synthesis and transport of secondary metabolites. *Proteomics* 12:3536-3547
- Chen Y, Xu Y, Luo W, Li W, Chen N, Zhang D, Chong K (2013) The F-Box protein OsFBK12 targets OsSAMS1 for degradation and affects pleiotropic phenotypes, including leaf senescence, in rice. *Plant Physiol* 163:1673 LP - 1685
- Dai Q, Wang H (2006) "Cullin 4 makes its mark on chromatin." *Cell Div* 1:1-7
- De Filippis L, Magel E (2012) Identification of biochemical differences between the sapwood and transition zone in *Robinia pseudoacacia* L. by differential display of proteins. *Holzforchung* 66:543-549
- Franciosini A, Lombardi B, Iafrate S, Pecce V, Mele G, Lupacchini L, Rinaldi G, Kondou Y, Gusmaroli G, Aki S, Tsuge T, Deng X-W, Matsui M, Vittorioso P, Costantino P, Serino G (2013) The *Arabidopsis* COP9 signalosome interacting F-box Kelch 1 protein forms an scf ubiquitin ligase and regulates hypocotyl elongation. *Mol Plant* 6:1616-1629
- Galindo-González L, Mhiri C, Deyholos MK, Grandbastien MA (2017) LTR-retrotransposons in plants: Engines of evolution. *Gene* 626:14-25
- Guo L, Nezames CD, Sheng L, Deng X, Wei N (2013) Cullin-RING ubiquitin ligase family in plant abiotic stress pathways. *J Integr Plant Biol* 55:21-30
- Harashima H, Schnittger A (2010) The integration of cell division, growth and differentiation. *Curr Opin Plant Biol* 13:66-74
- Hu X-Q, Han W, Han Z-Z, Liu Q-X, Xu X-K, Fu P, Li H-L (2014) A new macrocyclic lactone and a new quinoflavan from *Celastrus hindsii*. *Phytochem Lett* 7:169-172
- Jacobs DI, Gaspari M, Van Der Greef J, Van Der Heijden R, Verpoorte R (2005) Proteome analysis of the medicinal plant *Catharanthus roseus*. *Planta* 221:690-704
- Jia L, Yan F, Cao W, Chen Z, Zheng H, Li H, Pan Y, Narula N, Ren X, Li H, Zhou P (2017) Dysregulation of CUL4A and CUL4B ubiquitin ligases in lung cancer. *J Biol Chem* 292:2966-2978
- Jiang X, Hou H, Zhang S, Liu Y, Wang H, Deng WW, Zhou S, Wu Y, Shen F, Gao L, Xia T (2017) Comparison of phenolic compound accumulation profiles in eight evergreen woody core eudicots indicating the diverse ecological adaptability of *Camellia sinensis*. *Sci Hortic (Amsterdam)* 219:200-206
- Kumar M, Padula MP, Davey P, Pernice M, Jiang Z, Sablok G, Contreras-Porcia L, Ralph PJ (2017) Proteome analysis reveals extensive light stress-response reprogramming in the seagrass *Zostera muelleri* (Alismatales, Zosteraceae) metabolism. *Front Plant Sci* 7:2023
- Li C-L, Yang W-Z, Shi Z, Yuan HS (2018) Tudor staphylococcal nuclease is a structure-specific ribonuclease that degrades RNA at unstructured regions during microRNA decay. *Rna* 064501.117
- Li E, Ling J, Wang G, Xiao J, Yang Y, Mao Z, Wang X, Xie B (2015) Comparative proteomics analyses of two races of *Fusarium oxysporum* f. sp. *conglutinans* that differ in pathogenicity. *Sci Rep* 5:13663
- Li M, Tang D, Wang K, Wu X, Lu L, Yu H, Gu M, Yan C, Cheng Z (2011) Mutations in the F-box gene LARGER PANICLE improve the panicle architecture and enhance the grain yield in rice. *Plant Biotechnol J* 9:1002-1013
- Liu CL, Yang PS, Chien MN, Chang YC (2018) Expression of serine peptidase inhibitor Kunitz type 1 in differentiated thyroid cancer. *Histochem Cell Bio* 149:635-644
- Ly TN, Shimoyamada M, Yamauchi R (2006) Isolation and characterization of rosmarinic acid oligomers in *Celastrus hindsii* Benth. leaves and their antioxidative activity. *J Agric Food Chem* 54:3786-93
- Macedo MLR, Ribeiro SFF, Taveira GB, Gomes VM, Barros KMCA, Maria-Neto S (2016) Antimicrobial activity of ILTI, a Kunitz-type trypsin inhibitor from *Inga laurina* (SW.)

- Willd. *Curr Microbiol* 72:538-544
- Mandal A, Mishra AK, Dulani P, Muthamilarasan M, Shweta S, Prasad M (2018) Identification, characterization, 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. *Funct Integr Genomics* 18:101-111
- Manna S (2015) An overview of pentatricopeptide repeat proteins and their applications. *Biochimie* 113:93-99
- Matsunaga W, Ohama N, Tanabe N, Masuta Y, Masuda S, Mitani N, Yamaguchi-Shinozaki K, Ma JF, Kato A, Ito H (2015) A small RNA mediated regulation of a stress-activated retrotransposon and the tissue specific transposition during the reproductive period in *Arabidopsis*. *Front Plant Sci* 6:1-12
- Miyazaki Y, Yoshizumi T, Takase T, Matsui M, Kiyosue T (2011) Overexpression of *LOV KELCH PROTEIN 2* enhances cell elongation and increases cell number and ploidy in the hypocotyl of *Arabidopsis thaliana*. *Plant Biotechnol* 28:267-272
- Pokharel D, Padula MP, Lu JF, Jaiswal R, Djordjevic SP, Bebawy M (2016) The role of CD44 and ERM proteins in expression and functionality of P-glycoprotein in breast cancer cells. *Molecules* 21: 290
- Price SJ, Pangloli P, Krishnan HB, Dia VP (2016) Kunitz trypsin inhibitor in addition to Bowman-Birk inhibitor influence stability of lunasin against pepsin-pancreatin hydrolysis. *Food Res Int* 90:205-215
- Rocheta M, Carvalho L, Viegas W, Morais-Cecilio L (2012) Corky, a gypsy-like retrotransposon is differentially transcribed in *Quercus suber* tissues. *BMC Res Notes* 5:1-6
- Roodbarkelari F, Bramsiepe J, Weindl C, Marquardt S, Novak B, Jakoby MJ, Lechner E, Genschik P, Schnittger A (2010) CULLIN 4-RING FINGER-LIGASE plays a key role in the control of endoreplication cycles in *Arabidopsis* trichomes. *Proc Natl Acad Sci* 107:15275-15280
- Roy UK, Lavignac N, Rahman AM, Nielsen BV (2018) Purification of lectin and Kunitz trypsin inhibitor from soya seeds. *J Chromatogr Sci* 56:36-442
- Singh N, Sharma A (2017) Turmeric (*Curcuma longa*): miRNAs and their regulating targets are involved in development and secondary metabolite pathways. *Comptes Rendus - Biol* 340:481-491
- Singh N, Srivastava S, Sharma A (2016) Identification and analysis of miRNAs and their targets in ginger using bioinformatics approach. *Gene* 575:570-576
- Skaar I, Adaku C, Jordheim M, Byamukama R, Kiremire B, Andersen ØM (2014) Purple anthocyanin colouration on lower (abaxial) leaf surface of *Hemigraphis colorata* (Acanthaceae). *Phytochemistry* 105:141-146
- Thuy TT, Cuong NH, Sung T Van (2007) Triterpenes from *Celastrus hindsii* Benth. *J Chem* 373-376
- Vanderschuren H, Lentz E, Zainuddin I, Gruijssem W (2013) Proteomics of model and crop plant species: Status, current limitations and strategic advances for crop improvement. *J Proteomics* 93:5-19
- Viet TD, Xuan TD, Van TM, Andriana Y, Rayee R (2019) GC-MS and ESI-MS fingerprints of *Celastrus*. *Medicines* 6:1-15
- Voronova A, Belevich V, Jansons A, Rungis D (2014) Stress-induced transcriptional activation of retrotransposon-like sequences in the Scots pine (*Pinus sylvestris* L.) genome. *Tree Genet Genomes* 10:937-951
- Wang L, Pan D, Li J, Tan F, Hoffmann-Benning S, Liang W, Chen W (2015) Proteomic analysis of changes in the *Kandelia candel* chloroplast proteins reveals pathways associated with salt tolerance. *Plant Sci* 231:159-172
- Wang W, Lu JJ, Gu C, Zhou L, Liu S (2013) Performing isoelectric focusing and simultaneous fractionation of proteins on a rotary valve followed by sodium dodecyl-polyacrylamide gel electrophoresis. *Anal Chem* 85:6603-7
- Wang X, Li X, Deng X, Han H, Shi W, Li Y (2007) A protein extraction method compatible with proteomic analysis for the euhalophyte *Salicornia europaea*. *Electrophoresis* 28:3976-87
- Wen PF, Ji W, Gao MY, Niu TQ, Xing YF, Niu XY (2015) Accumulation of flavanols and expression of leucoanthocyanidin reductase induced by postharvest UV-C irradiation in grape berry. *Genet Mol Res* 14:7687-7695
- Wilkins MR, Williams KL (1997) Cross-species protein identification using amino acid composition, peptide mass fingerprinting, isoelectric point and molecular mass: a theoretical evaluation. *J Theor Biol* 186:7-15
- Witwer KW, Hirschi K.D (201) Transfer and functional consequences of dietary microRNAs invertebrates: Concepts in search of corroboration. *BioEssays* 36:394-406
- Wu SR, Teng CH, Tu YT, Ko CJ, Cheng TS, Lan SW, Lin HY, Lee MS (2017) The Kunitz domain I of hepatocyte growth factor activator inhibitor-2 inhibits matrix metalloproteinase activity and invasive ability of human prostate cancer cells. *Sci Rep* 7:1-19
- Xing A, Pan L, Gao J (2018) p100 functions as a metastasis activator and is targeted by tumour suppressing microRNA-320a in lung cancer. *Thorac Cancer*, 9:152-158
- Zhang J, Xiao J, Li Y, Su B, Xu H, Shan X, Song C, Xie J, Li R, Sharwood R (2017) PDM3, a pentatricopeptide repeat-containing protein, affects chloroplast development. *J Exp Bot* 267:27-37
- Zhang K, Novak O, Wei Z, Gou M, Zhang X, Yu Y, Yang H, Cai Y, Strnad M, Liu C-J (2014) *Arabidopsis* ABCG14 protein controls the acropetal translocation of root-synthesized cytokinins. *Nat Commun* 5:3274
- Zsigmond L, Rigo G, Szarka A, Szekely G, Otvos K, Darula Z, Medzihradsky KF, Koncz C, Koncz Z, Szabados L (2008) *Arabidopsis* PPR40 connects abiotic stress responses to mitochondrial electron transport. *Plant Physiol* 146:1721-1737