WALLACE, EMILY D., M.S. Identification of Adulteration in Botanical Samples with Untargeted Mass Spectrometry Metabolomics. (2019)
Directed by Dr. Nadja B. Cech. 62 pp.

According to the 2017 Council for Responsible Nutrition survey, botanicals makeup 39% of the total dietary supplement usage in the United States. The use of dietary supplements in general has increased by 8% since 2015, and there is a need to ascertain and monitor the quality and authenticity of such products. Adulteration of dietary supplements is a concern because commercial suppliers may inadvertently or deliberately sell products for which composition does not match that reported on the label. Adulteration constitutes a potential health concern for consumers, increasing the risk of toxicity, adverse reactions, or ineffective products. Current methodologies employ targeted analysis or supervised statistical analysis for adulteration detection, both of which require prior knowledge of the sample set. A method for detection of adulteration in botanical dietary supplements utilizing untargeted mass spectrometry based metabolomics was developed and implemented on multiple instrument platforms. These included an ultraperformance liquid chromatography (UPLC) system coupled to a ultraviolet (UV) spectrophotometric detector (LC-UV), UPLC coupled to a quadrupole-time of flight (Q-ToF) mass spectrometer (LC-Q-ToF) and UPLC coupled to a hybrid quadrupole-orbitrap mass spectrometer (LC-Orbitrap). To evaluate the sensitivity of the method for detecting outliers, a set of samples was prepared by combining two different plant species, the botanical *Hydrastis canadensis* L. (Ranunculaceae), and a known adulterant species, *Coptis chinensis* Franch. (Ranunculaceae). *C. chinensis* was added to

the *H. canadensis* samples in percentages ranging from 5% to 95% to emulate different levels of adulteration. The methodology was effective on all instrument platforms, but the sensitivity of detecting the adulterants varied depending on the analytical method and the method of data analysis. Using an unsupervised technique for data analysis (principal component analysis), the lowest percentage at which the adulterated sample was detectable as an outlier was measured based on the Hotelling's T^2 95% confidence interval. Outliers could be detected with this approach at 50%, 50%, and 10% adulteration using the LC-UV, LC-Q-ToF, and LC-Orbitrap systems, respectively. Composite score analysis was also performed for a statistical analysis comparison. A targeted analysis of a characteristic marker of adulteration (the alkaloid palmatine, which is a component of *C. chinensis*) was also conducted for comparison to the untargeted methods. Supervised statistical analyses, soft independent modelling by class analogy (SIMCA), was used to compare the sensitivity of different statistical approaches. The lowest percentage of adulteration detected as an outlier by these methods was 5%. SIMCA may be able to detect a lower percentage of adulteration, however, 5% was the lowest percentage tested in this study. The targeted analysis gave a limit of detection (LOD) of 0.0047 µM, 0.025 µM, and 0.027 µM; and a limit of quantitation (LOQ) of 0.12 µM and 0.55 µM, and 0.54 µM using liquid chromatography coupled to Orbitrap MS, Q-ToF MS, and Photodiode array (PDA) detectors, respectively. These values correspond to 0.3%, 1.5%, and 1.7% *C. chinensis* contamination in a botanical sample, respectively. Thus, a targeted methodology would detect trace levels of adulteration much more effectively than an untargeted method. However, untargeted methods have

the added advantage of being applicable even when the identity of adulterants is unknown.

IDENTIFICATION OF ADULTERATION IN BOTANICAL SAMPLES WITH UNTARGETED MASS SPECTROMETRY METABOLOMICS

by

Emily D. Wallace

A Thesis Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Greensboro
2019

Approved by

_____
Committee Chair

To my husband, Colby, for his continued support and reassurance, and my parents, for always encouraging my aspirations. I could not have done this without you.

APPROVAL PAGE

This thesis, written by EMILY D. WALLACE, has been approved by the following committee of the Faculty of the Graduate School at the University of North Carolina at Greensboro.

Committee Chair _____

Committee Members _____

_____

_____
Date of Acceptance by Committee

_____
Date of Final Oral Examination

ACKNOWLEDGMENTS

I have become the scientist that I am today because of the wonderful people that I have had the opportunity to work with. By name, I must thank Dr. Daniel Todd and Dr. Josh Kellogg for their wonderful mentorship, guidance, and knowledge. I would also like to thank Dr. Nadja B. Cech for motivating and helping me become a stronger person and chemist. Without her none of this work would be possible.

I owe acknowledgments to many more people. My dear friends, Victoria Meadows and Jennifer Simpson, for their continued support. My chemistry sisters, Emily Britton and Lindsay Caesar, for their support and friendship. The Cech lab as a whole, for being my second family and providing unrelenting support. Dr. Nicholas Oberlies, for his wise words and thoughtful input. The faculty and staff in the department for creating a cordial environment.

The National Institutes of Health National Center for Complementary and Integrative Health (NIH NCCIH), specifically the Center of Excellence for Natural Product Drug Interaction Research (NaPDI,U54AT008909) for providing financial support for this project.

TABLE OF CONTENTS

Page

LIST OF TABLES vi

LIST OF FIGURES vii

CHAPTER

I. INTRODUCTION 1

II. METHODS 5

General Methods 5
Sample Selection and Reference Materials 5
Sample Adulteration 6
Sample Extraction 7
Compound Identification 7
Mass Spectrometry Analysis 8
Metabolomics Analysis 9
Quantitative Analysis 10
LC-UV Metabolomics 11
Supervised Statistical Analysis 12

III. RESULTS AND DISCUSSION 13

Adulteration of Goldenseal Samples 13
Unsupervised Statistical Analysis 15
Quantitative Comparison 23
LC-UV Metabolomics 25
Composite Score Analysis 28
Supervised Statistical Analysis 31

IV. CONCLUSION 34

REFERENCES 38

APPENDIX A. SUPPLEMENTARY DATA 42

LIST OF TABLES

Page

Table 1. Composition of Commercial Botanical Products 6

Table 2. Composition of Adulterated Supplements 7

Table 3. Limit of Detection and Limit of Quantitation of Palmatine.....24

LIST OF FIGURES

Page

Figure 1. Extracted Ion Chromatograms of Three Adulterated Samples 14

Figure 2. Principal Component Analysis (PCA) Scores Plot from the Orbitrap (A) and Q-ToF (B) Data 17

Figure 3. PCA Loadings Plot for Orbitrap (A) and Q-ToF (B) Data 19

Figure 4. Outlier Limitation of Orbitrap (A) and Q-ToF (B)..... 22

Figure 5. LC-UV Scores and Loadings Plots..... 25

Figure 6. Composite Score Analysis of the Entire Sample Set (Tables 1 and 2)..... 29

Figure 7. Composite Score Analysis with 25% Adulterated Product 31

Figure 8. SIMCA-PCA Plot Constructed using All Goldenseal Supplements 33

CHAPTER I

INTRODUCTION

The 2017 Council for Responsible Nutrition survey found that botanicals makeup 39% of the total dietary supplement usage in the United States, the general use of which has increased by 8% since 2015[1]. Botanical dietary supplements encompass a wide range of over-the-counter products including capsules, tea, tinctures, and loose powders prepared from plant material[2]. The Dietary Supplement Health and Education Act (DSHEA) of 1994 assigns the Federal Drug Administration (FDA) regulatory oversight of dietary supplements[3]. Regulation and quality control of these products is challenging due to their inherent complexity and variability, and because the landscape of companies producing them is vast and constantly changing[4-5]. These regulatory and analytical challenges constitute a problem because contaminated or adulterated product may put the consumer at risk of adverse interactions[6].

Adulteration of a botanical occurs when the composition reported on the label does not match the actual material being sold[7]. This problem can occur due to limited availability of the natural product either via cultivation or ethical and legal wildcrafting, economic incentives of substituting other natural products or synthetic materials, or poor quality control during production[8-9]. A methodology using an untargeted metabolomics approach was recently developed to detect unknown adulteration in botanical dietary supplements[10].

One botanical product for which there is known to be a problem with contamination is *Hydrastis canadensis* L. (Ranunculaceae), commonly known as goldenseal. While the benzylisoquinoline alkaloid berberine is present in goldenseal and frequently attributed as the main bioactive principle, it is common across a wide variety of plants including *Berberis vulgaris* L. (Berberidaceae), *Mahonia aquifolium* (Pursh) Nutt. (Berberidaceae), and *Coptis chinensis* Franch. (Ranunculaceae)[11-12]. However, beyond berberine, these other species possess distinct metabolic profiles from that of goldenseal; two defining metabolites found in goldenseal are hydrastine and canadine, which are absent in other berberine-containing plants[7, 10, 13], while *B. vulgaris* (barberry), *M. aquifolium* (Oregon grape), and *C. chinensis* (Chinese goldthread) all have additional alkaloids including coptisine, dihydrocoptisine, palmatine, and jatorrhizine that are not present in goldenseal[14-16]. The presence of these marker compounds are signs of possible adulteration; however, trace amounts of contamination may or may not be detectable and there is often little to no prior knowledge of supplement composition or suspicion of adulteration.

There have been several studies conducted to assess the authenticity of goldenseal supplements, including targeted quantitative analysis, untargeted Fourier-transform near-infrared spectroscopy (FT-NIR) analysis, and untargeted ultra-performance liquid chromatography tandem mass spectrometry (UPLC-MS) metabolomics. Targeted analyses have the advantage of increased sensitivity and specificity compared to untargeted methodologies but require the identification of putative adulterants to be known prior to analysis. Several goldenseal studies employed targeted analysis using

HPLC-UV and GC-MS to detect and quantify metabolites in goldenseal commercial products[13, 17-18], and compounds from adulterating species were found in several of the commercial products[13].

Untargeted metabolomics methods have the benefit of comparing multiple complex products without any a priori knowledge of their composition or identification of major metabolites[7]. While it is not possible to measure the entirety of small molecules produced by an organism due to limitations in analytical methods, by detecting as many of these small molecules as possible, untargeted metabolomics approaches enable a holistic analysis in comparing complex samples[19-21]. Metabolomics has been utilized in a wide variety of applications including natural product drug discovery, dietary supplement adulteration, biological samples (cells, cancerous tissues, and fecal material)[10, 19], and botanical products (green tea, goldenseal, *Ginkgo biloba*, black cohosh, wheat, quinoa, and ginseng)[10, 19, 21-33]. In analyzing for potential adulteration, the variations in metabolomic profiles can represent alterations in the chemical composition, which could be attributed biological or genetic differences in the source material [27, 30-31, 34]. An untargeted metabolomics study using FT-NIR analysis assessed a sample set comprised of goldenseal and common adulterants[35]. In previous studies, goldenseal adulteration was simulated computationally for different adulterant species (yellow dock, yellow root, goldthread, and Oregon grape)[35]. Employing two supervised statistical analyses, SIMCA (soft independent modelling of class analogies) and PLS (partial least squares), a 5% adulteration level (e.g., 95% goldenseal, 5% adulterant) was identified as a statistical outlier[35]. However, computational spectra may or may not be indicative of collected

admixtures of spectra, especially since natural products are innately complex and variable.

In the current study, *H. canadensis* reference materials were physically (rather than computationally) mixed with *C. chinensis* plant material to form a series of intentionally adulterated products. These products will be analyzed using a UPLC-MS metabolomics method designed to detect adulteration in goldenseal products[10], while changing several analytical and statistical variables to compare approaches. Data will be acquired from multiple mass spectrometers, including a quadrupole-time of flight (Q-ToF) and an Orbitrap mass analyzer, to assess differences in sensitivity and method viability on different platforms. LC-UV data will be used as well for metabolomics input to show the method's potential on less sensitive analytical platforms. Composite score analysis will be utilized to compare different statistical analysis techniques[36]. SIMCA-PCA and PCA will be compared to show the effectiveness of supervised versus unsupervised analysis[37]. In addition, a targeted analysis of palmatine, a marker compound in berberine-producing plants that are common adulterants for *H. canadensis*, will be performed for comparison of targeted analysis to the untargeted methods. The methodology is expected to be effective for detecting statistical outliers on each platform applied, however, the range and sensitivity should vary based on the approach.

CHAPTER II

METHODS

General Methods

All solvents and chemicals used were of reagent or spectroscopic grade, as required, and obtained from ThermoFisher Scientific (Waltham, MA, USA) or Cayman Chemical (Ann Arbor, MI, USA). A palmatine chloride standard was purchased from Chromadex (Irvine, CA, USA) and was found to have a purity of 98% determined by UPLC-UV (data not shown).

Sample Selection and Reference Materials

Ten commercial goldenseal products were selected based on their popularity in online consumer sales reports[38]. All products were capsules and derived from root/rhizome material. Each sample was randomly coded with an internal reference number (beginning with the letters GS) to maintain manufacturer anonymity (Table 1).

Table 1. Composition of Commercial Botanical Products.

Sample Code	Composition	Form
GS-1	Root	Capsule
GS-2	Root	Capsule
GS-3	Root	Capsule
GS-4	Root	Capsule
GS-5	Root	Capsule
GS-6	Root	Capsule
GS-7	Root	Capsule
GS-8	Root	Capsule
GS-9	Root	Capsule
GS-10	Root/rhizome	Capsule
GS-11	Root	Capsule
GS-13	*Hydrastis canadensis* root	Loose Powder
GS-14	*Coptis chinensis* root	Loose Powder

GS-12 was eliminated from this study due to unsuspected adulteration, all other samples GS-1 through GS-11 were confirmed to be *Hydrastis canadensis* by profiling via mass spectrometery. GS-13 and GS-14 are the botanical reference material purchased from Chromadex for *Hydrastis canadensis* and *Coptis chineneis* respectively.

Botanical reference samples for *Hydrastis canadensis* root (GS-14) and *Coptis chinensis* root (GS-15) were obtained from Chromadex (Irvince, CA). Both reference materials were obtained as dried powders and extracted using the same methods applied for the *H. canadensis* samples.

Sample Adulteration

Samples were intentionally adulterated at different percentages by combining *H. canadensis* and *C. chinensis* at different ratios. A representative and verified *H. canadensis* commercial product, GS-4, was combined with the *C. chinensis* reference

material, GS-14, to achieve percentages of adulteration (Table 2). Samples were extracted as described below.

Table 2. Composition of Adulterated Supplements.

Sample Name	% w/w of *C. chinensis*	Mass *C. chinensis*	Mass *H. canadensis*
A-5	5%	10 mg	190 mg
A-10	10%	20 mg	180 mg
A-25	25%	50 mg	150 mg
A-50	50%	100 mg	100 mg
A-75	75%	150 mg	50 mg
A-90	90%	180 mg	20 mg
A-95	95%	190 mg	10 mg

Hydrastis canadensis and *Coptis chinensis* material were weighed out in different masses to arrive at a variety of percentages and a total mass of 200 mg. The percentage of *Coptis chinensis* is synonymous with the percentage of adulteration of the supplement.

Sample Extraction

Samples were weighed into scintillation vials (200 mg of material per sample) and 20.0mL of methanol were added. Extractions were performed in triplicate to provide process replicates for analysis. Samples were shaken for 24 hours then decanted into clean, weighed vials. Drying of extracts was accomplished under N_2 gas, and they were stored at room temperature prior to analysis.

Compound Identification

Variables, unique *m/z* value and retention time (*m/z*-RT) pairs, present in the loadings plot were used to confirm and explain the variance in the corresponding scores plot. These ions were identified by using exact mass (< 5 ppm) and retention time. The compounds (berberine [1], canadine [2], hydrastine [3], coptisine [4], palmatine [5],

jatrorrhizine [6], and dihydrocoptisine [7]) are all known and well documented (Supplemental, Figure S1). The *m/z*-RT pairs were compared and confirmed with literature values.

Mass Spectrometry Analysis

Two different mass spectrometers were employed. Liquid chromatography-mass spectrometry (LC-MS) data were acquired utilizing a Q Exactive Plus quadrupole-Orbitrap mass spectrometer (ThermoFisher Scientific) with an electrospray ionization (ESI) source coupled to an Acquity UPLC system (Waters, Milford, MA, USA). Samples were resuspended in methanol to a concentration of 0.1 mg/mL. Injections of 3 μL were performed on an Acquity UPLC BEH C_{18} column (1.7 μm, 2.1 × 50 mm, Waters) with a flow rate of 0.3 mL/min using the following binary solvent gradient of water (0.1% formic acid added) and acetonitrile (0.1% formic acid added): initial isocratic composition of 95:5 (water: acetonitrile) for 1.0 min, increasing linearly to 0:100 over 7 minutes, followed by an isocratic hold at 0:100 for 1 min, gradient returned to starting conditions of 95:5 and held isocratic again for 2 min. The positive ionization mode of the mass spectrometer was utilized over a full scan of *m/z* 150-900 with the following settings: capillary voltage, 5 V; capillary temperature, 300 °C; tube lens offset, 35 V; spray voltage, 3.80 kV; sheath gas flow and auxiliary gas flow, 35 and 20 units, respectively. Each sample was injected in triplicate to provide analytical replicates for analysis. Extracted ion chromatographs were obtained from the XCalibur software (ThermoFisher Scientific).

LC-MS data were also acquired using the Synapt G2 quadrupole-Time of Flight (Q-ToF) mass spectrometer (Waters) with an electrospray ionization (ESI) source coupled to an Acquity UPLC system. The same LC methodology and sample preparation were employed as described for the analysis on the Orbitrap system. The positive ionization mode of the mass spectrometer was utilized over a full scan of *m/z* 150−900 with the following settings: capillary voltage, 2.5 kV; sampling cone, -67 units; extraction cone, -106.2 units; source temperature, 80 °C; desolvation temperature, 150 °C; cone gas flow and desolvation gas flow, 1 L/h and 800 L/h, respectively. Each sample was injected in triplicate to obtain analytical replicates. Extracted ion chromatograms were obtained from the MassLynx software (Waters).

Metabolomic Analysis

The LC-MS data were analyzed, aligned, and filtered using MZmine 2.28 software (http://mzmine.github.io/) with a slightly modified version of a previously reported method[7]. The following parameters were used for peak detection of the data acquired from the Q Exactive Plus: noise level (absolute value), 1×10^5 counts; minimum peak duration 0.5 min; tolerance for *m/z* intensity variation, 20%. Peak list filtering and retention time alignment algorithms were performed to refine peak detection. The *join algorithm* was used to integrate all the chromatograms into a single data matrix using the following parameters: the balance between *m/z* and retention time was set at 10.0 each, *m/z* tolerance was set at 0.001 or 5 ppm, and retention time tolerance was defined as 0.5 min. The peak areas for individual ions detected in the process replicates and analytical replicates were exported from the data matrix for further analysis.

Data acquired from the Synapt G2 were analyzed in a similar fashion. Data were converted from project files (.PRO) to NetCDF files (.CDF) using Databridge in MassLynx (Waters, Milford, MA) and imported into MZmine for data processing. The following values were changed for processing the Q-ToF data:*m/z* tolerance was set at 0.01 or 10 ppm, and retention time tolerance was defined as 0.1 min. The peak areas for individual ions were exported from the data matrix for both the process and analytical replicates.

Relative standard deviation (RSD) filtering was utilized for all datasets. Analytical replicates would be expected to have comparable profiles, and similar peak areas for each feature[39]. Ions detected within the analytical replicates with disparate peak areas (based on an RSD cutoff of 25%), were assigned as artefacts of the instrument and excluded from the metabolomics analysis. The peak area of any feature (*m/z* and retention time pair) with an RSD value above 25% was replaced with a 0. Principal component analysis (PCA) was performed using Sirius version 10.0 (Pattern Recognition Systems AS, Bergen, Norway). Data transformation was carried out by a fourth root transform of peak area to reduce heteroscedasticity. The 95% confidence interval was calculated using Hoetelling's T^2 with the R package 'car' [40].

Quantitative Analysis

Targeted analysis was performed using a palmatine standard purchased from Chromadex. Standards were prepped at a range of concentrations in optima grade methanol (Supplemental, Table S2) using serial dilutions. Quantified extracts were prepped at a concentration of 0.1 mg/mL. The same parameters and LC method was utilized on all three

systems as stated previously. On the Orbitrap platform a selective ion monitoring (SIM) scan was performed from a range of 350.1549-354.1549 for heightened sensitivity. A similar approach was used on the Q-ToF with a *m/z* window of 350-354. The LC-UV data were collected in a range of 150-600 nm, but for processing purposes a range of 346.3-346.4 nm was selected out.

The limit of detection (LOD) for each approach was calculated using the following equation: $LOD = \frac{3S}{m}$ where s is the standard deviation of the lowest point in the linear range and m is the slope of the regression line[41]. The limit of quantitation (LOQ) was determined as the lowest concentration of standard in the calibration curve that provided a residual of less than 15% (formula: (measured concentration-theoretical concentration)/ theoretical concentration × 100%). The limit of detection was converted into two different forms, ppm palmatine in the plant and w/w % of *Coptis chinensis* adulterant, for comparison to the untargeted methodologies. The ppm palmatine in the plant was calculated using the initial plant mass (199.10 mg) and extract mass (45.78 mg) of the *C. chinensis* reference material. The w/w % of *C. chinensis* adulterant was calculated using the ppm of palmatine in the plant and the amount of palmatine in the *C. chinensis* reference material (1.24 mg of palmatine per gram of plant material)

LC-UV Metabolomics

LC-UV data were collected in the same run on the Q Exactive Plus mass spectrometer, using the photodiode array detector (PDA) on the Waters Acquity UPLC across a range of 189 - 600 nm. The retention time and peak area for each sample were exported from Xcalibur into Excel. A data matrix was created of all the samples with

retention time and peak area. This was analyzed with Sirius to produce the principal component analysis (PCA) scores and loadings plots.

Supervised Statistical Analysis

SIMCA-PCA analysis was conducted in Sirius version 10.0. A mean of objects normalization was applied to the Orbitrap data. The SIMCA model employed a training set of GS 1 through 11 (confirmed goldenseal products) and GS-13 (reference material) (Table 1). Four components were extracted from this model for a total of 94.0% variance explained. All samples were then fit to the model for comparison.

CHAPTER III

RESULTS AND DISCUSSION

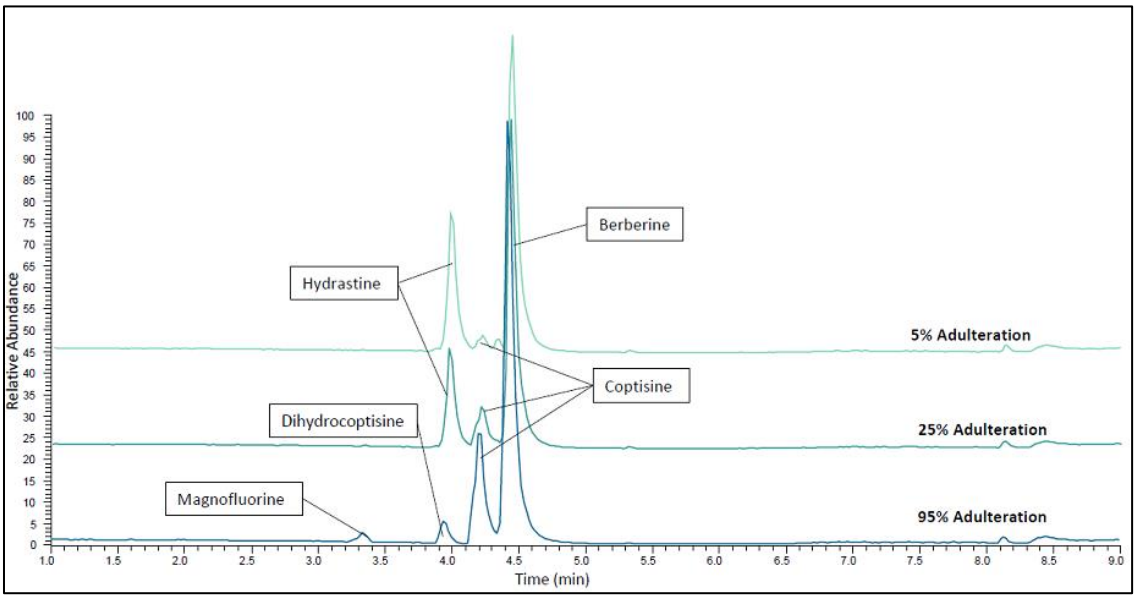
Adulteration of Goldenseal Samples

C. chinensis possesses characteristic marker compounds: palmatine, dihydrocoptisine, coptisine, and magnoflorine[16]. These compounds were found to increase in abundance, corresponding with an increase in *C. chinensis* material (Figure 1). There are several compounds that are representative of goldenseal that are absent in other berberine-containing species. Hydrastine and canadine are two abundant alkaloids that are unique to goldenseal[13, 42]. As the percentage of goldenseal decreased in the adulterated samples, the relative intensity of these alkaloids also decreased. A marked difference is not noticeable within the base peak chromatogram until 25% adulteration, where a distinct shift in the ratio between hydrastine and berberine was observed (Figure 1). Coptisine and dihydrocoptisine are visible in the base peak chromatogram level at this point as well. Figure 1 shows the appearance/disappearance of these characteristic ions across three different adulteration percentages on both the Orbitrap and Q-ToF mass spectrometers. Both sets of chromatograms show the same trend as the percentage of *C. chinensis* increases, however, the chromatograms from the Q-ToF show a change in relative abundance as well. A 22% increase in berberine signal was detected in the 95% adulterated sample compared to the 5% adulterated sample. This suggests a higher

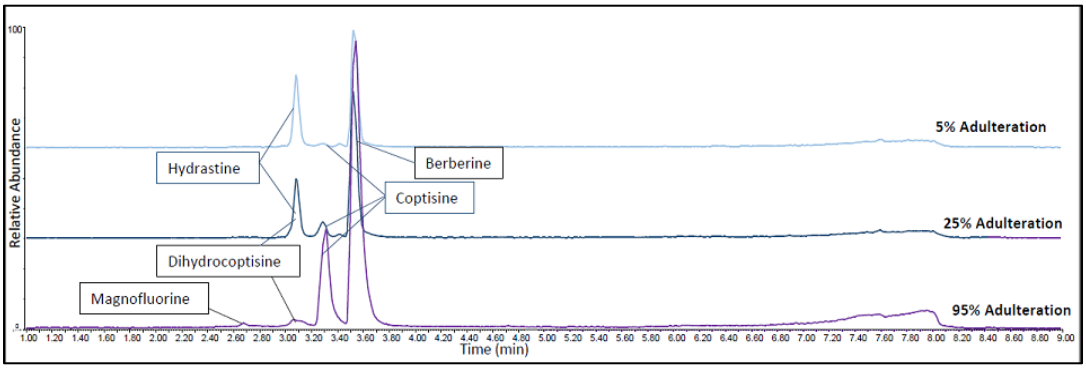
percentage of berberine in *C. chinensis* as compared to *H. canadensis.* No increase in berberine signal for the samples with higher *C. chinenesis* content was observed in the Orbitrap data, suggesting that the berberine signal was saturated in the Orbitrap data.

Figure 1. Extracted Ion Chromatograms of Three Adulterated Samples.

A



B



Each set includes chromatograms of 5, 25, and 95% adulteration (percentage of *Coptis chinensis* present, remaining material is *Hydrastis canadensis*). One set (A) is chromatograms from the Orbitrap (Q-Exactive); while the other (B) is from the Q-ToF (Synapt G2). Marker compounds were identified by accurate mass and retention time, <5 ppm for the Orbitrap and <10 ppm for the Q-ToF. Hydrastine ($[M+H]^+$ 384.1435) and canadine ($[M+H]^+$ 340.1538) are marker compounds in *Hydrastis canadensis*, while magnofluroine ($[M+]^+$ 342.1700), coptisine ($[M+]^+$ 320.0918), dihydrocoptisine ($[M+]^+$ 322.1075), and palmatine ($[M+]^+$ 352.1542) are unique to *Coptis chinensis*. As the percentage of *Coptis chinensis* in

the samples increases, the presence of hydrastine and canadine decreases and is replaced by an increase in coptisine, palmatine, magnofluroine, and dihydrocoptisine. Chromatograms were all normalized to the same relative intensity for comparison.

Unsupervised Statistical Analysis

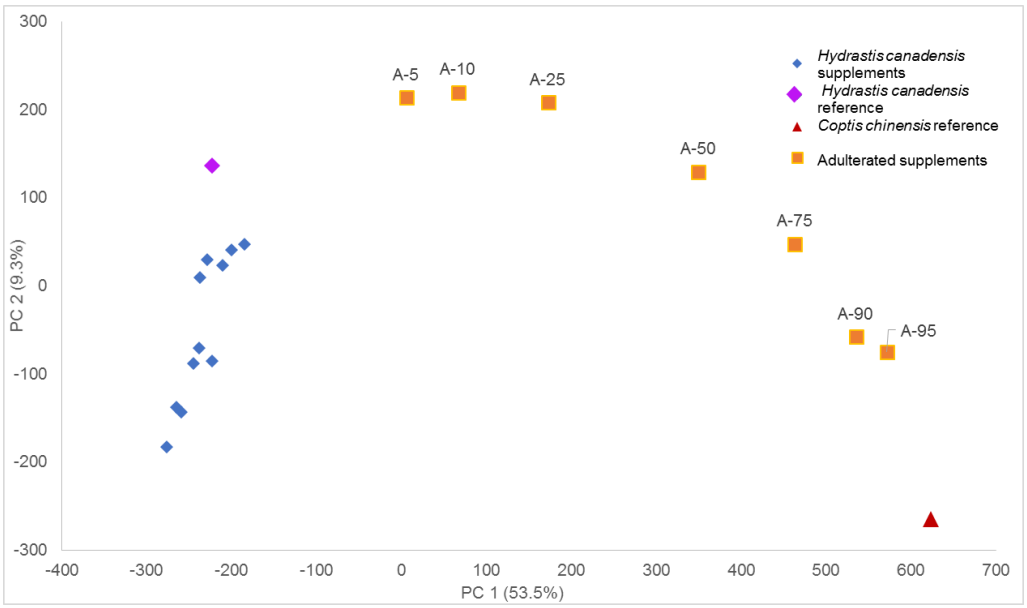
Untargeted metabolomics analysis using principal component analysis (PCA) was performed on three datasets to determine at which percentage adulteration could be detected. Datasets from two different mass spectrometers were compared in a qualitative way using untargeted metabolomics. The general profile of the samples were the same, as shown in Figure 1. The baseline of each dataset were set at approximately 40 ppm of the total intensity for a coverage of 5 orders of magnitude each for comparative purposes. A greater number of features was detected utilizing the mass spectrometer with an Orbitrap mass analyzer (Q Exactive Plus), approximately 4,000, while the mass spectrometer with a Q-ToF mass analyzer (Synapt G2) detected approximately 1,000. Several factors could affect the number of ions detected on each platform (ionization source settings, quadrupole mass filtering, scan range, etc.). These features may also be attributed to chemical interference within the LC system (the same gradient was used on both instruments but two separate LC systems were used) as well as background noise within the instrument itself. All of these factors may affect the metabolomics data output and consequently make it more or less sensitive.

A PCA scores plot shows the relationship between different samples, where each data point is representative of that sample's chemical profile (as described by features detected and associated abundance). The PCA data for each mass spectrometry platform (Figure 2) evidenced a similar trend in percentage of adulteration; i.e., the higher the

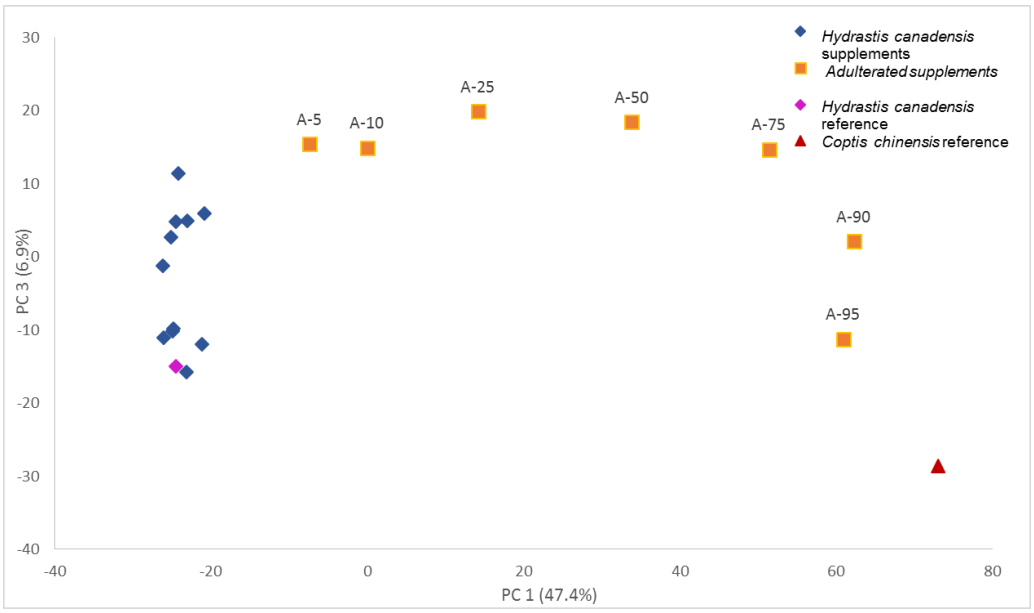
adulteration, the further that the adulterated sample (orange squares) was spatially from the cluster of unadulterated goldenseal samples (blue diamonds). The purple diamond and red triangle represent the goldenseal and *C. chinensis* reference materials, respectively. The goldenseal reference material clusters with the group of commercial supplements, while the *C. chinensis* reference is observed to cluster further away from the sample with 95% *C. chinensis* and 5% goldenseal. This suggests that the 95% adulterated supplement is not pure *C. chinensis*, rather, it still contains some constituents found in goldenseal. The same trend was observed on both instrument platforms, which shows the main compounds driving the untargeted metabolomics analysis are likely the more abundant compounds. This can be confirmed by looking at the loadings plot to see which variables (*m/z* and retention time) are responsible for the differences observed among samples.

Figure 2. Principal Component Analysis (PCA) Scores Plot from the Orbitrap (A) and Q-ToF (B) Data.

A



B



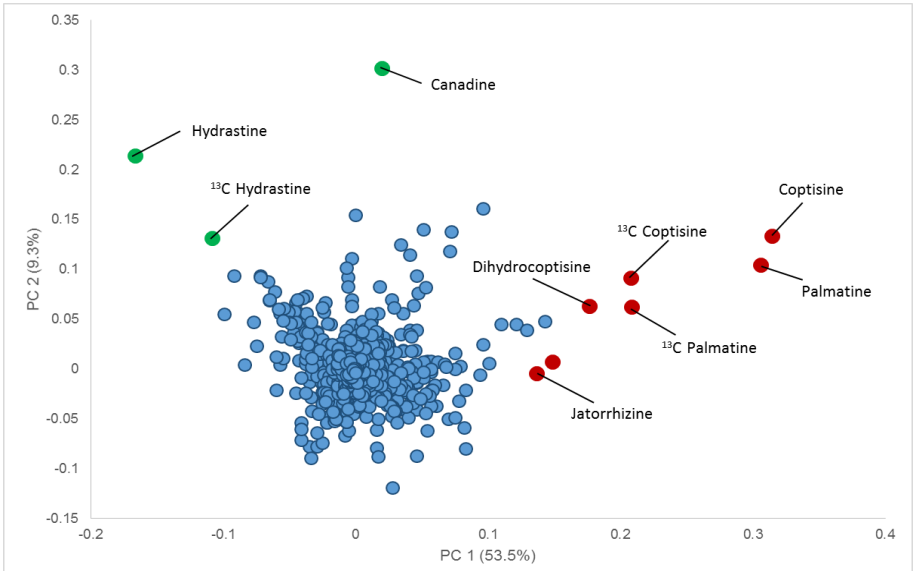
The points are labeled to indicate the percentage of adulterant (*Coptis chinensis*) added to the *Hydrastis canadensis* plant material (A-5 is 5% *C. chinensis*, 95% *H. canadensis*, A-10 is 10% *C. chinensis*, etc.). The same trend was observed in both scores plots, with adulterated samples shifted increasingly from the cluster of authentic *H. canadensis* samples as the percentage of adulterant increased. Similar variance was showcased with a PC 1 (53.5%) versus PC 2 (9.3%) comparison of the Orbitap data for a total of 62.8% variance, and a PC 1 (47.7%) versus PC 3 (11.5%) comparison for the Q-ToF for a total of 59.2%

explained variation. The red triangle in each plot denotes the reference material for *C. chinensis*, while the purple diamond represents the reference material for *H. canadensis*.

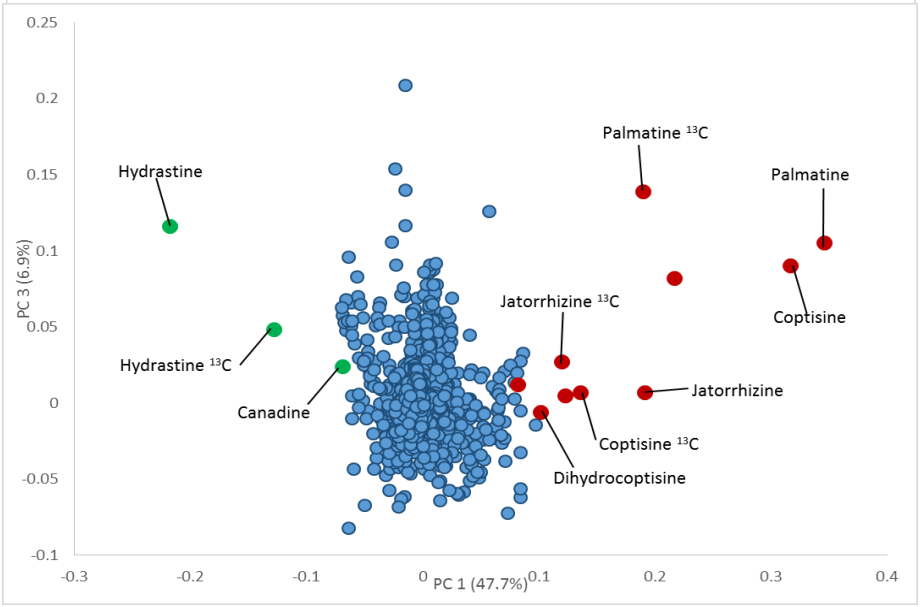
The corresponding loadings plots describe the relationship between the samples in the scores plot, by geographically plotting the variables (*m/z* and retention time pair) that are attributed to the variance (Figure 3). Thus, the loadings can be used to qualitatively correlate the unique variables to the representative samples in the scores plot. The loadings plot highlights the compounds that are known to be characteristic of each plant species; the green markers are compounds unique to goldenseal, while red markers represent metabolites unique to *C. chinensis*. Hydrastine, canadine, as well as the ^{13}C isotope of hydrastine are located in the upper left region, which corresponds to the position of the goldenseal supplements in the scores plot (Figure 2). Palmatine, coptisine, dihydrocoptisine, and the ^{13}C isotopes of palmatine and coptisine are visible in the lower right region of the plot, which corresponds to the position of the *C. chinensis* reference material and the adulterated samples (Figure 2). This supports the distinction between groupings of samples observed in the scores plot.

Figure 3. PCA Loadings Plot for Orbitrap (A) and Q-ToF (B) Data.

A



B



Each data point represents a variable, which is a unique *m/z* value and retention time pair detected in the sample. The variables in red are associated with *Coptis chinensis*, while those in green are associated with *Hydrastis canadensis*. Dihydrocoptisine was observed in the loadings plot for the Orbitrap data but not in the Q-ToF data. Conversely, more jatorrhizine associated adducts were observed in the Q-ToF plot but not the Orbitrap data.

The marked difference between the Orbitrap and the Q-ToF platforms were observed in the loadings plot in the detection of multiple ions corresponding to the

compound jatorrhizine, which is unique to *C. chinensis* but lower in abundance than major alkaloids palmatine, coptisine, and berberine[16]. The $[M+]^+$ value was detected in loadings plots from both instruments, however, there were additional ion derivatives in the loadings plot for the Q-ToF (Figure 3B) including the ^{13}C isotope peak. Thus, a larger amount of relevant ions were detected at a higher abundance using a Q-ToF mass analyzer, some of which were not identified (marked in red). This was probably due to the limitations of the Orbitrap analyzer, as the capacity of the trap is limited to a finite quantity of ions at a given time, described by the automatic gain control (AGC) target. Thus, the Orbitrap can become filled with higher abundant ions, suppressing lower abundant compounds eluting at a similar time.

The percentage of adulterant that would result in a sample being characterized as an outlier was determined for untargeted Principal Component Analysis (PCA) by using Hotelling's T^2 95% confidence interval. This calculation can be applied to multivariate data to assess similarity among samples- samples that fall within the confidence interval are believed to be similar with 95% certainty, while samples that lie beyond the confidence ellipse are considered statistically distinct (within 5% error). A confidence ellipse is a common statistical descriptor for determining significance and outliers in a sample set without needing prior knowledge of the chemical composition[10]. In the case where all adulterated and unadulterated samples were included (Tables 1 and 2), the application of the Hotelling's T^2 test enabled several outliers (the adulterated samples containing 75% through 95% *C. chinensis*) to be distinguished from the rest of the samples (Supplemental, Figure S2). However, there is an inherent challenge with the

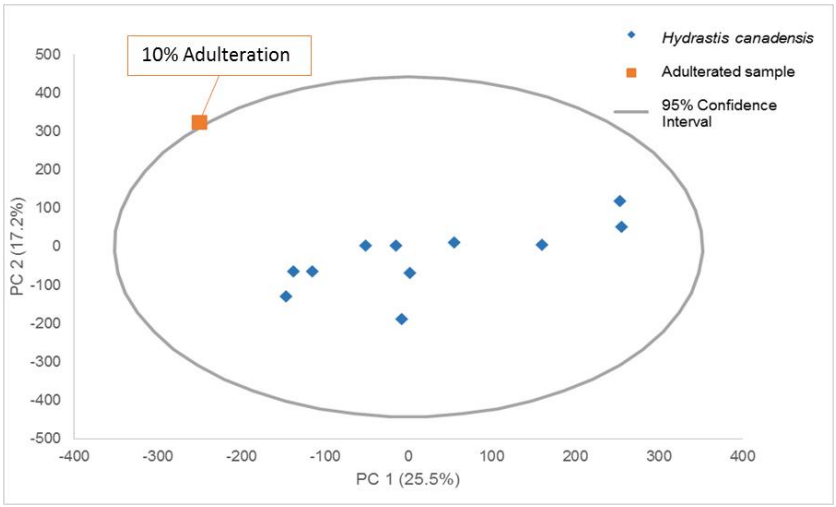
inclusion of many adulterated samples within the datasets, as the model that is generated is skewed towards adulterated samples and the ability to distinguish between adulterated and unadulterated samples is reduced. For comparison, when only one adulterated sample was included with the rest of the unadulterated goldenseal samples, it was possible to distinguish adulterated samples from unadulterated samples by the Hotelling's T^2 method at a much lower percentage of *C. chinensis* added (Figure 4).

The datasets obtained with different mass spectrometry platforms differed greatly in the percentage of adulterant that could be distinguished as an outlier. For the data obtained with the Orbitrap mass analyzer, the adulterated sample became a visible outlier (i.e., was spatially located beyond the confidence ellipse) at a 10% w/w *C. chinensis* composition (or 10% adulteration). The Q-ToF platform required 50% adulterant before the adulterated sample could be distinguished, far higher than the Orbitrap. One factor that may explain this is the differences in sensitivity of the Q-ToF and Orbitrap instruments; the sensitivity of the Q-ToF platform is generally greater than that of the Orbitrap, however, in this case it is clear that the Orbitrap platform data results in the most sensitive metabolomics output. This could take place for numerous reasons (ionization source parameters, optics, detector gain, etc.) and does not infer anything about the instrument's performance itself[43-44]. A higher sensitivity will allow a lower quantity of analyte to be detected with a higher signal output. If these lower-abundant ions contribute to the differences between adulterated and unadulterated samples, improved sensitivity will result in improved ability to detect an outlier. The base peak chromatograms obtained with the Q-ToF and Orbitrap mass analyzers look very similar

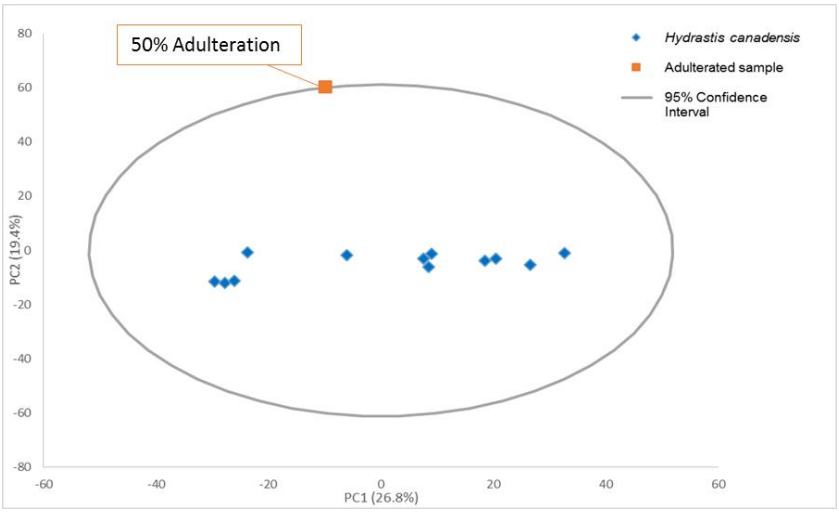
(Figure 1), as the same chromatographic gradient was applied in both cases. The instrument platforms themselves are quite different and multiple factors go into the data collected, however, the decrease in detectable adulteration suggests that the more sensitive the mass spectrometer the more sensitive the metabolomics analysis will be. Metabolomics can be utilized on a variety of mass spectrometry platforms, however, the Orbitrap platform yielded data that resulted in more effective metabolomics analysis.

Figure 4. Outlier Limitation of Orbitrap (A) and Q-ToF (B).

A



B



The Hotelling's T^2 95% confidence interval is shown as an elipse. The 95% confidence interval was applied to PCA analysis of the goldenseal samples with only a single adulterated sample included in the dataset. The inclusion of just one outlier improves the ability to detect outliers as compared to datasets where multiple outliers were included (Supplemental, Figure S2). Concerning the Orbitrap data, at 5% adulteration the sample was not an outlier however (Supplemental, Figure S3), at 10% adulteration the sample was beyond the confidence interval and became a visible outlier (A). The outlier limitation in a dataset that include just one outlier using the Orbitrap is determined to be 10%. The Q-ToF data was less effective for distinguishing outliers, and the sample with 25% adulteration was still not an observed outlier (Supplemental, Figure S3), but at 50% (B) the adulterated sample was observed outside the confidence ellipse.

Quantitative Comparison

Palmatine was quantified on both instruments to compare a targeted approach to the untargeted metabolomics analysis. A targeted approach (specifically selecting for the known alkaloid palmatine, present in the *C. chinensis* adulterant) was much more sensitive on all three platforms (Table 3) than the untargeted methods. For example, even the least sensitive of the three instrument platforms, the LC-UV system, had a limit of detection for palmatine of 0.027 µM. This corresponds to 20 ppm palmatine in a plant sample, or 1.7% *Coptis chinensis* adulterant. In other words, it would theoretically be possible to detect palmatine using targeted analysis by LC-UV in a sample of *H. canadensis* containing 1.7% or more *C. chinensis*. The mass spectrometric methods were

even more sensitive, with the Orbitrap system giving a theoretical limit of 0.3% *C. chinensis* (Table 3). These values are well below even the lowest cut-off (10% adulterant) observed with untargeted metabolomics using unsupervised data analysis. Thus, for situations where the adulterant is of known identity, a targeted analysis will be able to detect adulteration at much lower levels (33 fold in this case). However, it is worth noting that a disadvantage of targeted analysis is that it does require a priori knowledge as to the identity of the adulterant. If analysis was conducted on a sample and there was not any suspicion or prior knowledge concerning adulteration, or if the identity of potential marker compounds was not known, it would not be possible to utilize a targeted analysis.

Table 3. Limit of Detection and Limit of Quantitation of Palmatine.

Method of analysis	**Limit of detection (LOD) palmatine (µM)[a]**	**Limit of quantitation (LOQ) palmatine (µM)[b]**	**Limit of detection expressed as ppm palmatine in *C. chinensis* plant[c]**	**Minimum detectable *Coptis chinensis* (% w/w)[d]**
LC with UV	0.027	0.54	20	1.7
LC-MS (Orbitrap)	0.0047	0.12	3.8	0.30
LC-MS (ToF)	0.025	0.55	18	1.5

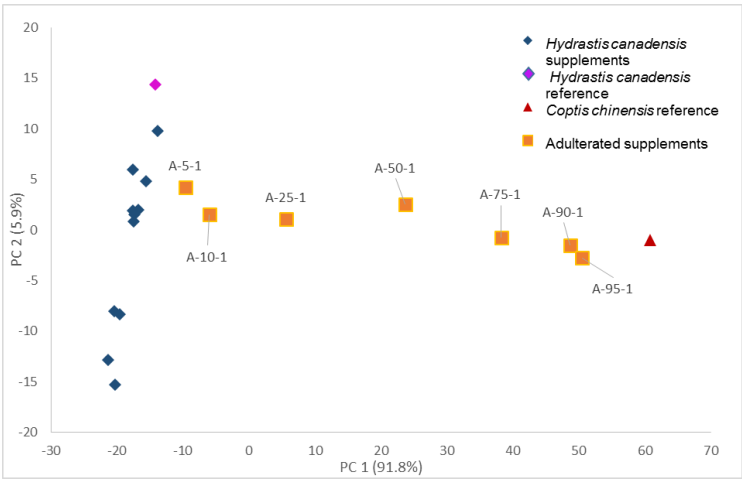
[a]Limit of detection was calculated using the following equation: $LOD = \frac{3S}{m}$ where s is the standard deviation and m is the slope from the regression line.
[b]Limit of quantitation was determined as the lowest concentration of standard in the calibration curve that provided a residual of less than 15%[45].
[c]Calculated using the limit of detection and the original plant mass and extract mass of the *Coptis chinensis* reference material to give a value of ppm palmatine in the plant.
[d]The w/w % of *Coptis chinensis* adulterant that would yield a concentration of palmatine corresponding to the limit of detection, calculated using the quantity of palmatine in the *Coptis chinensis* reference material (1.24 mg of palmatine per gram).

LC-UV Metabolomics

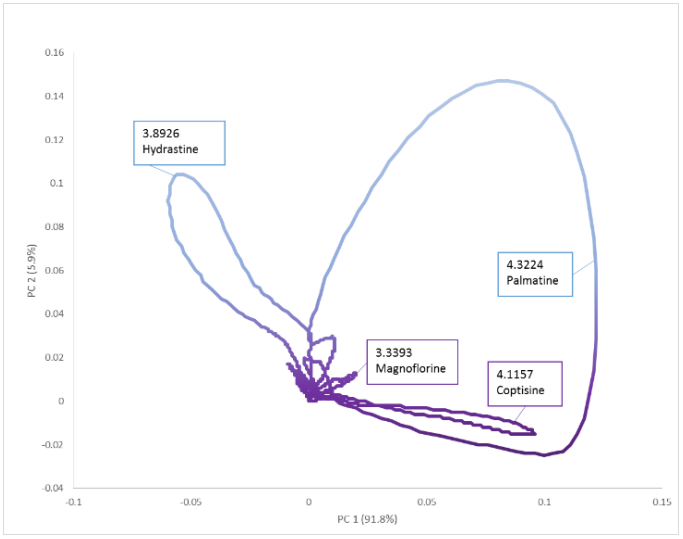
The data obtained using liquid chromatography separation with an ultraviolet (UV) photodiode array (PDA) detector were also utilized for metabolomics analysis. Using UV or PDA data as an input source has appeal as a more viable and cost-efficient option for entities that might not have access to mass spectrometry equipment. With UV/PDA data, the independent variables were only retention times, and peak intensity was used in place of peak area for PCA analysis (Figure 5).

Figure 5. LC-UV Scores and Loadings Plots.

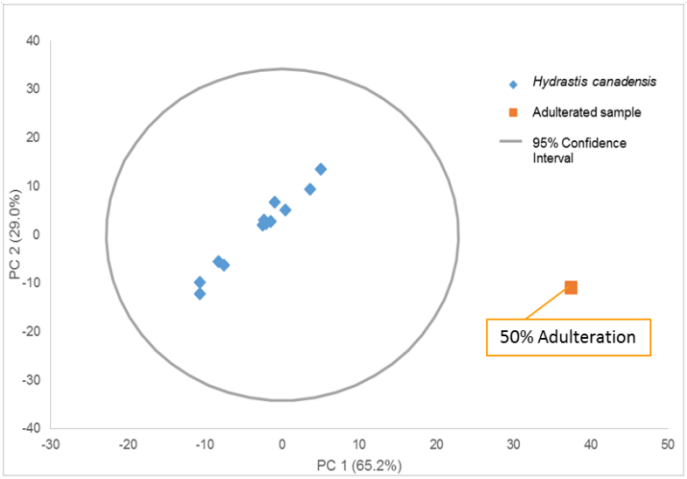
A



B



C



The scores plot (A) yielded a similar trend in adulteration percentage as seen with mass spectrometry data (Figure 2). The adulterated samples are represented as orange squares, while the goldenseal samples (Table 1) are represented by blue diamonds. The *C. chinensis* and *H. canadensis* references were spatially located with the appropriate clusters. The loadings plot (B) showed the retention time and directionally correlated with the scores plot. The same compounds as before (Figure 3) are shown to contribute to the variance, however, dihydrocoptisine and jatorrhizine are absent from the list of *C. chinensis* ions. Canadine is also absent from the list of *H. canadensis* compounds. An adulterated sample of 25% falls on the 95% confidence interval line (Supplemental, Figure S3). This may or may not be deemed an outlier depending on the situation. A sample with 50% adulteration (C) is a clear outlier, plotting far outside the 95% confidence interval.

The same trend in composition was observed here as well as in the mass spectrometry-based metabolomics (Figure 2). The variance of the adulterated samples are

proportional to the peak height of the unique metabolites in *C. chinensis*, as the peak height increases so does the weight of that variable within the statistical analysis. The PC 1 versus PC 2 scores plot encompassed 97.7% of the dataset's variance. UV absorbance is generally not as sensitive as mass spectrometry, however, this did not seem to impact the general PCA scores plot (Figure 5A). One main disadvantage of utilizing UV absorbance data for metabolomics analysis approach is the reduction in useful information gleaned from the loadings plot. With no discrete *m/z* values as input data, the loadings plot is a near-continuous plot of retention time (Figure 5B). Thus, the loops visible in the loadings plot correspond to the increase in intensity associated with peak formation at a given retention time. However, using only the UV absorbance data yielded little additional information to discern responsible metabolites underpinning the visible trends. Mass spectrometry provides additional information, such as *m/z* value to improve identification of unknown compounds and determine which metabolites are responsible for the variation observed in the samples. LC-UV metabolomics was successfully able to establish one of the adulterated samples as a tentative outlier as well (Figure 5C), with 50% w/w adulteration falling beyond the 95% confidence ellipse, and was comparable to the results obtained with the Q-ToF mass spectrometer (Figure 4B).

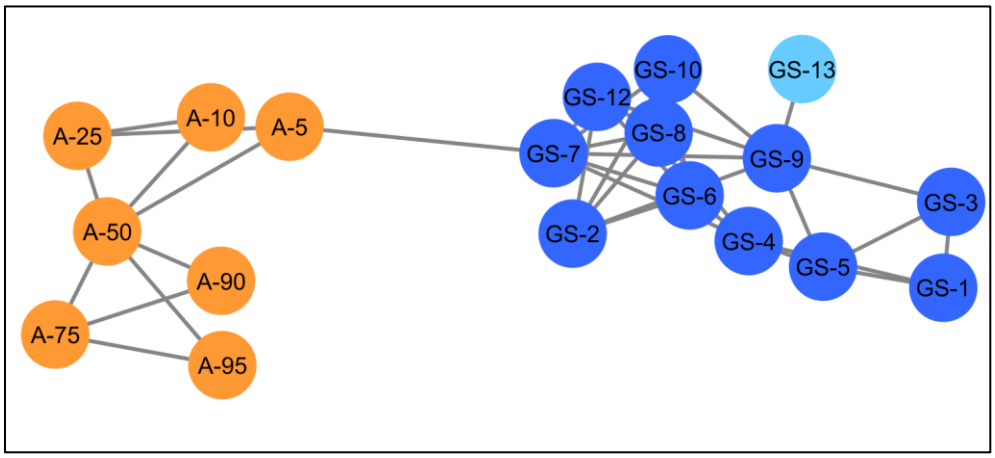
The Orbitrap mass spectrometer (Q Exactive) was able to detect adulterated samples as outliers at the lowest percentage (10% w/w adulteration), while the Q-Tof (Synapt G2) was 50% and the PDA required between 25 and 50% adulterant. The advantage of collecting mass spectrometry data was apparent in the additional information garnered about the metabolites, chemical profile, and being able to relate the

chemical composition to the variation of the samples. Pairing *m/z* value with retention time allowed for putative identification of secondary metabolites, as well as identification of possible adulteration sources, in this case *C. chinensis*. However, the LC-UV metabolomics approach is a more cost-effective analytical input to gauge sample relationships and authenticity, and could be improved if more was known about the sample set. Ultimately, all three dataset sources and both untargeted and targeted methodologies were successful in detecting adulterated samples, however, the Q-ToF and LC-UV were not as sensitive as the Orbitrap platform and may not be as useful in application.

Composite Score Analysis

Composite score analysis was performed using the data acquired on the Orbitrap mass analyzer (Q Exactive). In the previous approaches, the PCA data was limited to just two principal components (in these cases, PC1 and PC2). However, there is additional variation in the dataset that is not encapsulated within only two principal components. Expanding the analysis to include multiple principal components allows for a more comprehensive analysis of the dataset.

Figure 6. Composite Score Analysis of the Entire Sample Set (Tables 1 and 2).



The adulterated samples, in orange, are separated from the goldenseal commercial samples (in blue). The light blue node represents the goldenseal vouchered reference material. The connected lines represent a similarity score of >0.3, the highest score being 1.0.

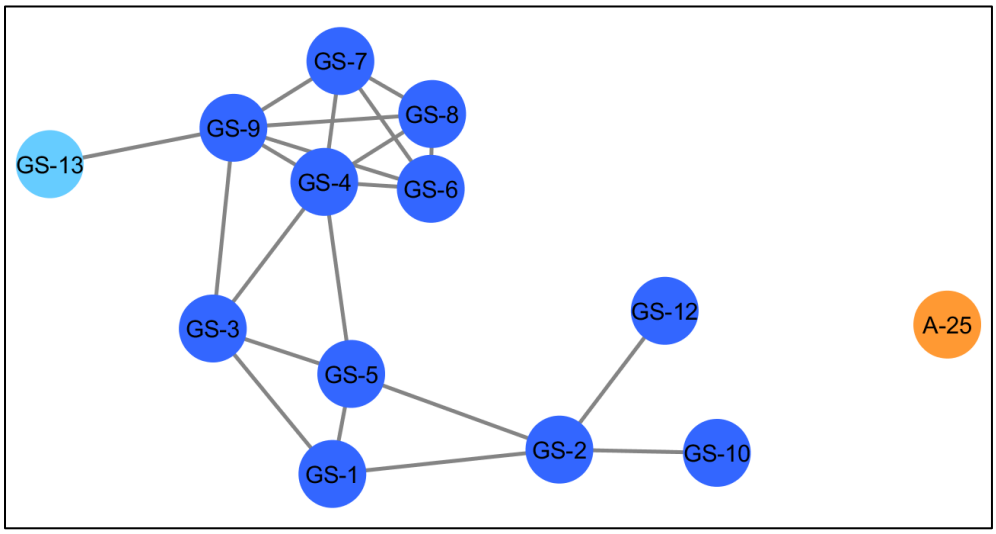
From the multi-component PCA, a similarity score was calculated between every sample, a correlation coefficient that ranges from -1.0 to 1.0. The correlations can serve as the foundation for a network diagram, with nodes (individual samples) while connections are derived from the correlation to connect nodes. For the analysis of all samples (Figure 7) a similarity score threshold of >0.3 was set. From the composite score analysis, there are two distinct clusters were observed: the adulterated samples (orange) and the unadulterated goldenseal samples (blue) (Tables 1 and 2). More connections were observed in the composite score analysis plot with all positive connections (0-1.0) but the two groups were still distinct (Supplemental, Figure S6).

When calculating a composite score network comparing the goldenseal sample cluster against a single adulterated product (Figure 8), the analysis was not as sensitive as principal component analysis based upon the Orbitrap untargeted metabolomics (Figure 4A). This is due to the similarity between the plants; there should be some overlap in

metabolite content given that the plants belong to the same family (Ranunculaceae). In addition, the "adulterated" samples are still partially comprised of *H. canadensis* so a few connections should be expected. By restricting the connectivity to a threshold of 0.3, the distinction between the two groups is clear but it reduces the overall picture of the model. However, using the composite score's network diagram facilitates visual determination of potential outlier samples; the similarity score serves as a quantitative measure to differentiate dissimilar samples. The cutoff point is relative and will vary among datasets and combination of samples, however, it provides an important metric for authentication.

In this sample set, using a similarity score of >0.10, 25% *Coptis chinensis* was completely differentiable from the goldenseal sample cluster with no connections. Again, this is a more restrictive way to use this model compared to a general similarity comparison but it works successfully in this application. This approach would be more sensitive if the same product were being analyzed rather than an assortment or if the two samples were vastly different botanicals. However, composite score analysis is a useful way to use and display principal component analysis data in a more quantitative and comprehensive way than a traditional 1x1 principal component comparison.

Figure 7. Composite Score Analysis with 25% Adulterated Product.



With a similarity score range of 0.10-1.0 the 25% adulterated product (orange) is no longer connected to the other nodes (blue).

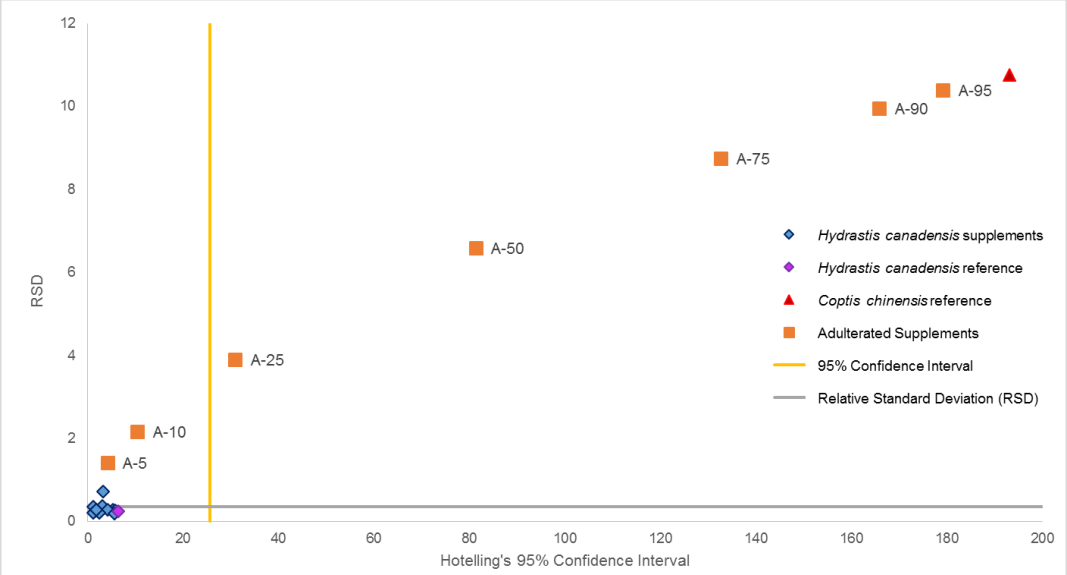
Supervised Statistical Analysis

The metabolomics analysis shown earlier was completed using PCA, which is an unsupervised statistical analysis. To compare supervised and unsupervised statistics, SIMCA-PCA was utilized. SIMCA-PCA builds a model based on a subset of reference samples, samples that should be indicative of the group (Figure 8). The model was conducted using the *H. canadensis* supplements and *H. canadensis* reference material as a training set (Table 1). All of the samples, adulterated and unadulterated, are then fit to this model for comparison. The 95% confidence interval (yellow line) and relative standard deviation (grey line) were calculated using the model built on the *H. canadensis* supplements. The RSD value can also be phrased as the distance of the sample from the model. If the sample is vastly different from the model it will have a much higher RSD value. This is shown as the composition of *C. chinensis* increases in the adulterated

samples. The Hotelling's T^2 95% confidence interval is used similarly to the application earlier, however, here it is calculated based on the model. Since the model is built upon actual botanical samples, the RSD value and 95% confidence interval are more indicative of the botanical and allow for a more cohesive comparison of the dataset. Thus, it is imperative to construct an appropriate and representative model in order to yield an accurate comparison of the dataset.

Only one *H. canadensis* supplement (GS-1) is an outlier, indicated by the relative standard deviation cutoff (grey line); however, with the Hotelling's T^2 95% confidence interval as an outlier indicator none of the *H. canadensis* supplements are outliers. After inspection of what could cause this differentiation from the rest of the *H. canadensis* supplements (blue diamonds), this particular product has less hydrastine than the majority of products. The peak area of hydrastine in GS-1, 3.95×10^9, is similar to the area of the 5% adulterated sample, 4.10×10^9, in comparison to the *H. canadensis* reference material, 5.04×10^9 (Supplemental, Figure S7). However, since the SIMCA model is more representative of the samples--the relative standard deviation could be used for outlier detection in place of the Hotelling's T^2 95% confidence interval. If a sample is plotted outside the RSD cutoff then it could be marked as suspicious and a follow up targeted analysis could be performed to see if the product is actually adulterated. This makes this approach much more sensitive with a detectable amount of adulteration of 5% rather than the 25% detectable with the 95% confidence interval.

Figure 8. SIMCA-PCA Plot Constructed using All Goldenseal Supplements.



The 95% confidence interval is indicated by the yellow line (value=25.67) and the grey line represents the relative standard deviation (RSD) cutoff (value=0.35). Both of these values were calculated using the model.

CHAPTER IV

CONCLUSION

There is a critical need to authenticate natural product-based dietary supplements due to the legal ambiguity in which the product evaluation and reporting of adverse effects remains the responsibility of the manufacturer[3]. When adulteration or risk of illness or injury are presented the burden of proof is placed upon the FDA[3]. Consumers are generally not well-informed of the potential health implications of taking adulterated botanical supplements. Adulterating products with the intent to maintain the presence of a bioactive principle (such as substituting *C. chinensis* for *H. canadensis* in a supplement) is not without risks, as bioactivity is a complex function. Multiple synergists, additives, and antagonists may be present in one natural product and may modulate the total activity. Thus, substituting with the aim of preserving one active compound - such as berberine - does not mean two different species are comparable in bioactivity, and other adverse effects could also manifest, including allergenicity or even toxicology[46].

This methodology provided an untargeted process for ascertaining the authentication of supplements as well as a targeted methodology that results in quantification of marker compounds. Untargeted metabolomics can be used as a tool to identify adulterated samples, and provide information about potentially unknown marker compounds that contribute to the differentiation, which is especially beneficial in

situations when there is little prior knowledge of the composition or adulterant. Targeted analysis can be used for a direct and quantitative comparison in conjunction to verify the level of adulteration present.

Untargeted metabolomics was able to discern authentic and adulterated *H. canadensis* samples using data obtained from a variety of analytical platforms. Two of the most popular forms of high-resolution mass spectrometry instrumentation, Q-ToF and Orbitrap platforms, both yielded datasets capable of discriminating between samples. The sensitivity was varied, with the lowest detectable percentage of adulteration for the Q-ToF being 50% compared to 10% for the Orbitrap. Employing UV absorbance as an analytical input yielded a dataset that was as sensitive as the Q-ToF, with a detectable amount of 50% adulteration. The targeted analysis was the most sensitive approach with low limits of detection (0.0047 µM, 0.033 µM, and 0.25 µM for the Orbitrap, Q-ToF, and PDA respectively) on all platforms. It is worth noting that neither of these methods were optimized: a general chromatography gradient method was used on both instruments with non-optimal separation (e.g, berberine and palmatine coeluted) and the ionization was also not optimized (a general tune file suitable for 300 µL LC flow was utilized). The UV/PDA method was also not optimized and was used in a very general form. If any of these elements were optimized the sensitivity and overall performance could have improved. However, using a more general method for this test case demonstrated that all three platforms were viable for detection of botanical adulteration. Another limitation of this study was the range of adulteration tested (5-95%). A few samples at lower

percentages (1-2%) would have provided a more accurate representation of how effective each statistical analysis and platform was at detecting adulteration.

Different statistical techniques may be applied to make this approach more sensitive. The unsupervised principal component analysis, a dominant statistical analytical method applied to metabolomics datasets[47], and yielded a more narrow view of the dataset and produced less sensitive quality control measures. However, PCA was able to discriminate between the different sample classes and provide illuminating statistical comparisons for the data. Composite score analysis had a similar sensitivity but combined four principal components to encompass a larger percentage of the variation in the dataset, as compared to the two principal component comparisons of traditional PCA. Supervised statistical analysis, SIMCA-PCA, was also used. SIMCA-PCA created a more comprehensive model based on the *H. canadensis* products and thus allowed for a more sensitive comparison. SIMCA-PCA was the most sensitive approach since the RSD value could be used for outlier detection. This gave a detectable amount of 5% adulteration. The disadvantage of model-based statistics is the relationship between the samples is not further related to the variables. PCA describes the correlation of the samples by using the variables directionally in the loadings plot. This is helpful for follow up targeted analysis or for identifying the potential adulterants. While each of these statistical techniques can be successfully applied for authentication of goldenseal samples, model-based statistics (e.g., SIMCA) will yield a more sensitive quality control measure if reference samples are available.

In this study, different commercial products were used to provide a more robust test case to challenge the analytical and statistical methods. In other settings, such as an industrial quality control environment, an increased number of authenticated products would increase the sensitivity of any of the statistical methodology, as more references or authenticated materials would tighten the variation among goldenseal samples and heighten the variance between the outlier and the goldenseal product clusters. In situations where a mass spectrometer is not accessible, LC-UV metabolomics offers a more affordable but comparable option. Regardless of the analytical instrumentation, untargeted metabolomics for adulteration detection could be adapted and enhanced for implementation in various applications.

REFERENCES

1. Vogtman, H. Dietary Supplement Usage Incrases, Says New Survey. https://www.crnusa.org/newsroom/dietary-supplement-usage-increases-says-new-survey (accessed November 5th).
2. Brown, A. C., An overview of herb and dietary supplement efficacy, safety and government regulations in the United States with suggested improvements. Part 1 of 5 series. *Food and chemical toxicology : an international journal published for the British Industrial Biological Research Association* **2017**, *107* (Pt A), 449-471.
3. Dietary Supplement Health and Education Act. In *103-417*, **1994**.
4. Newman, D. J.; Cragg, G. M., Natural Products as Sources of New Drugs from 1981 to 2014. *Journal of natural products* **2016**, *79* (3), 629-61.
5. Dwyer, J. T.; Coates, P. M.; Smith, M. J., Dietary Supplements: Regulatory Challenges and Research Resources. *Nutrients* **2018**, *10* (1).
6. Izzo, A. A.; Hoon-Kim, S.; Radhakrishnan, R.; Williamson, E. M., A Critical Approach to Evaluating Clinical Efficacy, Adverse Events and Drug Interactions of Herbal Remedies. *Phytotherapy research : PTR* **2016**, *30* (5), 691-700.
7. Tims, M. On Adulteration of *Hydrastis canadensis* root and rhizome Botanical Adulterants Bulletin [Online], **2016**.
8. McGraw JB, S. S., Van der Voort M, Distribution and abundance of *Hydrastis canadensis* L. (Ranunculaceae) and *Panax quinquefolius* L. (Araliaceae) in the central Appalachian region. *J Torrey Bot Soc* **2003**, *130* (2), 62-69.
9. Lee, J., Marketplace analysis demonstrates quality control standards needed for black raspberry dietary supplements. *Plant foods for human nutrition* (Dordrecht, Netherlands) **2014**, *69* (2), 161-7.
10. Wallace, E. D.; Oberlies, N. H.; Cech, N. B.; Kellogg, J. J., Detection of adulteration in *Hydrastis canadensis* (goldenseal) dietary supplements via untargeted mass spectrometry-based metabolomics. *Food and chemical toxicology : an international journal published for the British Industrial Biological Research Association* **2018**, *120*, 439-447.
11. Pengelly, A.; Bennett, K.; Spelman, K.; Tims, M., An Appalachian Plant Monograph: Goldenseal *Hydrastis canadensis* L. *Appalachian Center for Ethnobotanical Studies* **2012**.
12. Cicero, A. F. G.; Baggioni, A., Berberine and Its Role in Chronic Disease. *Advances in experimental medicine and biology* **2016**, *928*, 27-45.
13. Weber, H. A.; Zart, M. K.; Hodges, A. E.; Molloy, H. M.; O'Brien, B. M.; Moody, L. A.; Clark, A. P.; Harris, R. K.; Overstreet, J. D.; Smith, C. S., Chemical comparison of goldenseal (*Hydrastis canadensis* L.) root powder from three commercial suppliers. *Journal of agricultural and food chemistry* **2003**, *51* (25), 7352-8.

14. Ivanovska, N.; Philipov, S., Study on the anti-inflammatory action of *Berberis vulgaris* root extract, alkaloid fractions and pure alkaloids. *International journal of immunopharmacology* **1996**, *18* (10), 553-61.
15. Rackova, L.; Majekova, M.; Kost'alova, D.; Stefek, M., Antiradical and antioxidant activities of alkaloids isolated from *Mahonia aquifolium*. Structural aspects. *Bioorganic & medicinal chemistry* **2004**, *12* (17), 4709-15
16. Yang, Y.; Peng, J.; Li, F.; Liu, X.; Deng, M.; Wu, H., Determination of Alkaloid Contents in Various Tissues of *Coptis chinensis* Franch. by Reversed Phase-High Performance Liquid Chromatography and Ultraviolet Spectrophotometry. *Journal of chromatographic science* **2017**, *55* (5), 556-563.
17. Avula, B.; Wang, Y.-H.; Khan, I. A., Quantitative determination of alkaloids from roots of *Hydrastis canadensis* L. and dietary supplements using ultra-performance liquid chromatography with UV detection. *Journal of AOAC International* **2012**, *95* (5), 1398-405.
18. Brown, P. N.; Roman, M. C., Determination of hydrastine and berberine in goldenseal raw materials, extracts, and dietary supplements by high-performance liquid chromatography with UV: collaborative study. *Journal of AOAC International* **2008**, *91* (4), 694-701.
19. Harnly, J.; Chen, P.; Sun, J.; Huang, H.; Colson, K. L.; Yuk, J.; McCoy, J.-A. H.; Reynaud, D. T. H.; Harrington, P. B.; Fletcher, E. J., Comparison of Flow Injection MS, NMR, and DNA Sequencing: Methods for Identification and Authentication of Black Cohosh (*Actaea racemosa*). *Planta medica* **2016**, *82* (3), 250-62.
20. Hong, E.; Lee, S. Y.; Jeong, J. Y.; Park, J. M.; Kim, B. H.; Kwon, K.; Chun, H. S., Modern analytical methods for the detection of food fraud and adulteration by food category. *Journal of the science of food and agriculture* **2017**, *97* (12), 3877-3896.
21. Rodriguez, S. D.; Rolandelli, G.; Buera, M. P., Detection of quinoa flour adulteration by means of FT-MIR spectroscopy combined with chemometric methods. *Food chemistry* **2019**, *274*, 392-401.
22. Britton, E. R.; Kellogg, J. J.; Kvalheim, O. M.; Cech, N. B., Biochemometrics to Identify Synergists and Additives from Botanical Medicines: A Case Study with *Hydrastis canadensis* (Goldenseal). *Journal of natural products* **2017**.
23. Caesar, L. K.; Kellogg, J. J.; Kvalheim, O. M.; Cech, R. A.; Cech, N. B., Integration of Biochemometrics and Molecular Networking to Identify Antimicrobials in *Angelica keiskei*. *Planta medica* **2018**, *84* (9-10), 721-728.
24. Deconinck, E.; Sokeng Djiogo, C. A.; Courselle, P., Chemometrics and chromatographic fingerprints to classify plant food supplements according to the content of regulated plants. *Journal of pharmaceutical and biomedical analysis* **2017**, *143*, 48-55.
25. Geng, P.; Harnly, J. M.; Chen, P., Differentiation of Whole Grain from Refined Wheat (*T. aestivum*) Flour Using Lipid Profile of Wheat Bran, Germ, and Endosperm with UHPLC-HRAM Mass Spectrometry. *Journal of agricultural and food chemistry* **2015**, *63* (27), 6189-211.
26. Geng, P.; Harnly, J. M.; Sun, J.; Zhang, M.; Chen, P., Feruloyl dopamine-O-hexosides are efficient marker compounds as orthogonal validation for authentication of

black cohosh (*Actaea racemosa*)-an UHPLC-HRAM-MS chemometrics study. *Analytical and bioanalytical chemistry* **2017**, *409* (10), 2591-2600.
27. Kang, J.; Lee, S.; Kang, S.; Kwon, H. N.; Park, J. H.; Kwon, S. W.; Park, S., NMR-based metabolomics approach for the differentiation of ginseng (*Panax ginseng*) roots from different origins. *Archives of pharmacal research* **2008**, *31* (3), 330-6.
28. Karu, N.; Deng, L.; Slae, M.; Guo, A. C.; Sajed, T.; Huynh, H.; Wine, E.; Wishart, D. S., A review on human fecal metabolomics: Methods, applications and the human fecal metabolome database. *Analytica chimica acta* **2018**, *1030*, 1-24.
29. Kellogg, J. J.; Graf, T. N.; Paine, M. F.; McCune, J. S.; Kvalheim, O. M.; Oberlies, N. H.; Cech, N. B., Comparison of Metabolomics Approaches for Evaluating the Variability of Complex Botanical Preparations: Green Tea (*Camellia sinensis*) as a Case Study. *Journal of natural products* **2017**, *80* (5), 1457-1466.
30. Kortesniemi, M.; Sinkkonen, J.; Yang, B.; Kallio, H., NMR metabolomics demonstrates phenotypic plasticity of sea buckthorn (*Hippophae rhamnoides*) berries with respect to growth conditions in Finland and Canada. *Food chemistry* **2017**, *219*, 139-147.
31. McGeachie, M. J.; Dahlin, A.; Qiu, W.; Croteau-Chonka, D. C.; Savage, J.; Wu, A. C.; Wan, E. S.; Sordillo, J. E.; Al-Garawi, A.; Martinez, F. D.; Strunk, R. C.; Lemanske, R. F., Jr.; Liu, A. H.; Raby, B. A.; Weiss, S.; Clish, C. B.; Lasky-Su, J. A., The metabolomics of asthma control: a promising link between genetics and disease. *Immunity, inflammation and disease* **2015**, *3* (3), 224-38.
32. Pinasseau, L.; Vallverdu-Queralt, A.; Verbaere, A.; Roques, M.; Meudec, E.; Le Cunff, L.; Peros, J.-P.; Ageorges, A.; Sommerer, N.; Boulet, J.-C.; Terrier, N.; Cheynier, V., Cultivar Diversity of Grape Skin Polyphenol Composition and Changes in Response to Drought Investigated by LC-MS Based Metabolomics. *Frontiers in plant science* **2017**, *8*, 1826.
33. van Gorsel, M.; Elia, I.; Fendt, S.-M., 13C Tracer Analysis and Metabolomics in 3D Cultured Cancer Cells. *Methods in molecular biology* (Clifton, N.J.) **2019**, *1862*, 53-66.
34. Ghatak, A.; Chaturvedi, P.; Weckwerth, W., Metabolomics in Plant Stress Physiology. *Advances in biochemical engineering/biotechnology* **2018**.
35. Liu, Y.; Finley, J.; Betz, J. M.; Brown, P. N., FT-NIR characterization with chemometric analyses to differentiate goldenseal from common adulterants. *Fitoterapia* **2018**, *127*, 81-88.
36. Kellogg, J. J.; Paine, M. F.; McCune, J. S.; Oberlies, N. H.; Cech, N. B., Selection and characterization of botanical natural products for research studies: a NaPDI center recommended approach. *Natural product reports* **2019**.
37. Vitale, R.; Marini, F.; Ruckebusch, C., SIMCA Modeling for Overlapping Classes: Fixed or Optimized Decision Threshold? *Analytical chemistry* **2018**, *90* (18), 10738-10747.
38. Amazon.com Amazon Best Sellers. https://www.amazon.com/Best-Sellers-Health-Personal-Care-Goldenseal-Herbal-Supplements/zgbs/hpc/3765771/ref=zg_bs_nav_hpc_3_3764461 (accessed January **2017**).

39. Caesar, L. K.; Kvalheim, O. M.; Cech, N. B., Hierarchical cluster analysis of technical replicates to identify interferents in untargeted mass spectrometry metabolomics. *Analytica chimica acta* **2018**, *1021*, 69-77.
40. Fox J, W. S., *An R Companion to Applied Regression*. 2nd ed.; Sage: Thousand Oaks, CA, 2011.
41. Harris, D. C., *Quantitative Chemical Analysis*. W. H. Freeman and Company: New York, United States, 2002.
42. Le, P. M.; McCooeye, M.; Windust, A., Characterization of the alkaloids in goldenseal (*Hydrastis canadensis*) root by high resolution Orbitrap LC-MS(n). *Analytical and bioanalytical chemistry* **2013**, *405* (13), 4487-98.
43. Haag, A. M., Mass Analyzers and Mass Spectrometers. *Advances in experimental medicine and biology* **2016**, *919*, 157-169.
44. Eliuk, S.; Makarov, A., Evolution of Orbitrap Mass Spectrometry Instrumentation. *Annual review of analytical chemistry* (Palo Alto, Calif.) **2015**, *8*, 61-80.
45. Todd, D. A.; Zich, D. B.; Ettefagh, K. A.; Kavanaugh, J. S.; Horswill, A. R.; Cech, N. B., Hybrid Quadrupole-Orbitrap mass spectrometry for quantitative measurement of quorum sensing inhibition. *Journal of microbiological methods* **2016**, *127*, 89-94.
46. Wu, K.-M.; Ghantous, H.; Birnkrant, D. B., Current regulatory toxicology perspectives on the development of herbal medicines to prescription drug products in the United States. *Food and chemical toxicology : an international journal published for the British Industrial Biological Research Association* **2008**, *46* (8), 2606-10.
47. Bartel, J.; Krumsiek, J.; Theis, F. J., Statistical methods for the analysis of high-throughput metabolomics data. *Computational and structural biotechnology journal* **2013**, *4*, e201301009.

APPENDIX A

SUPPLEMENTARY DATA

Figure S1. Structures of Key Compounds in *Hydrastis canadensis* and *Coptis chinensis*

Table S1. *m/z* Values of Key Compounds and the Species Associated

Table S2. Concentrations of Palmatine used for Calibration Curves

Figure S2. PCA Scores Plot Showing All Samples with the 95% Confidence Interval

Figure S3. Highest Percentage of Adulterated Sample that was not an Outlier on all Platforms

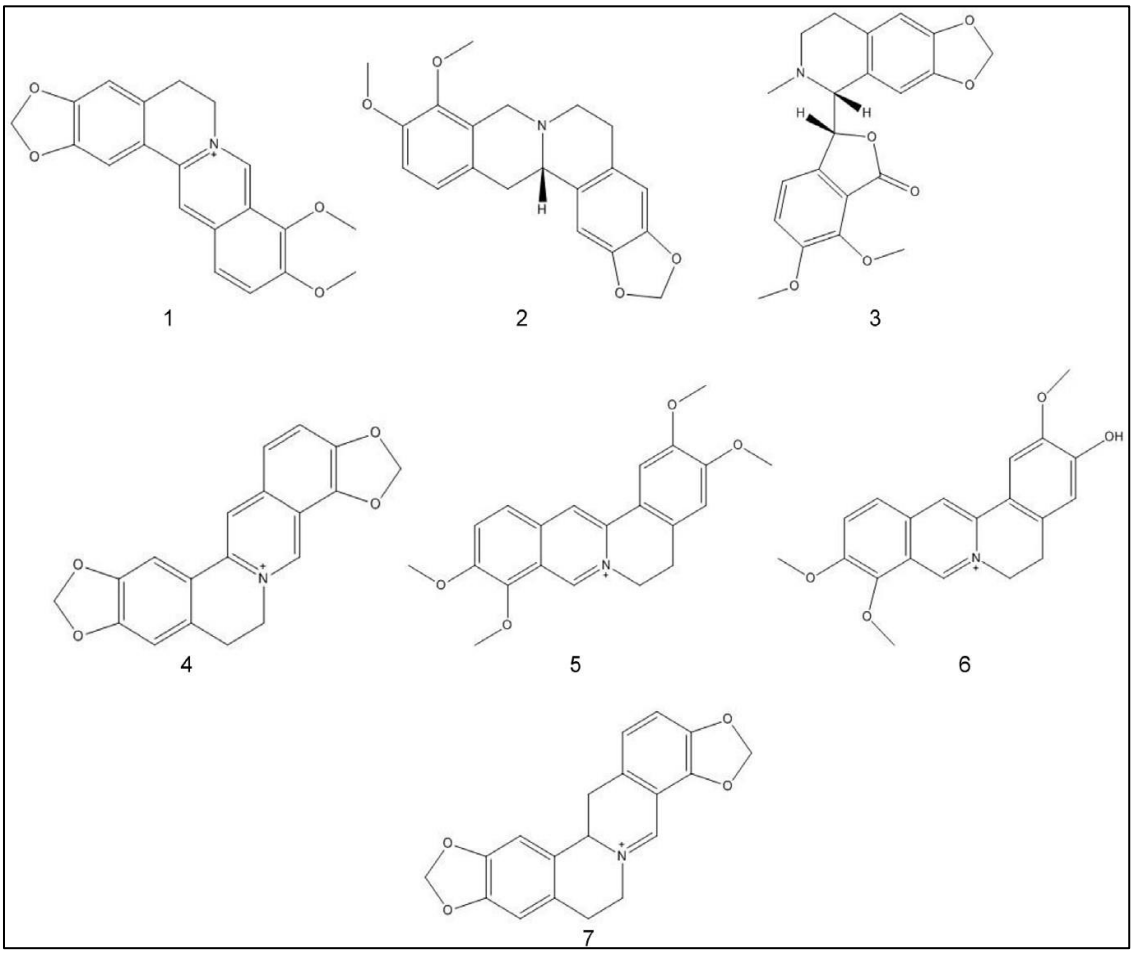
Figure S4. Stacked LC-UV Total Absorbance Chromatograms

Figure S5. Calibration Curves for Palmatine

Figure S6. Composite Score Analysis Plot with all Positive Connections

Figure S7. Peak Integration of Hydrastine

Figure S1. Structures of Key Compounds in *Hydrastis canadensis* and *Coptis chinensis.*



These compounds were confirmed using exact mass and retention time. The compounds are berberine (1), canadine (2), hydrastine (3), coptisine (4), palmatine (5), jatrorrhizine (6), and 13,14 dihydrocoptisine (7).

Table S1. *m/z* Values of Key Compounds and the Species Associated.

Compound Name	***m/z* Value**	**Species Associated**
Berberine	336.1229	*Hydrastis canadensis, Coptis chinensis*
Hydrastine	384.1440	*Hydrastis candensis*
Canadine	340.1545	*Hydrastis canadensis*
Sideroxylin	313.1066	*Hydrastis canadensis*
Coptisine	320.0916	*Coptis chinensis*
Dihydrocoptisine	322.1074	*Coptis chinensis*
Palmatine	352.1543	*Coptis chinensis*
Jatorrhizine	338.1392	*Coptis chinensis*

Table S2. Concentrations of Palmatine used for Calibration Curves.

LC-UV (µg/mL)	LC-MS (Orbitrap (µg/mL)	LC-MS (Q-ToF) (µg/mL)
0.048*	0.0025	0.048*
0.097	0.0050	0.097
0.19	0.012	0.19
0.39	0.025	0.39
1.5	0.048	1.5
3.1	0.097	3.1
6.2	0.19	6.2
12.5	0.39	12.5
25*	0.78	25*

*Concentration not included in calibration curve due to being outside the linear dynamic range.

Figure S2. PCA Scores Plot Showing All Samples with the 95% Confidence Interval.

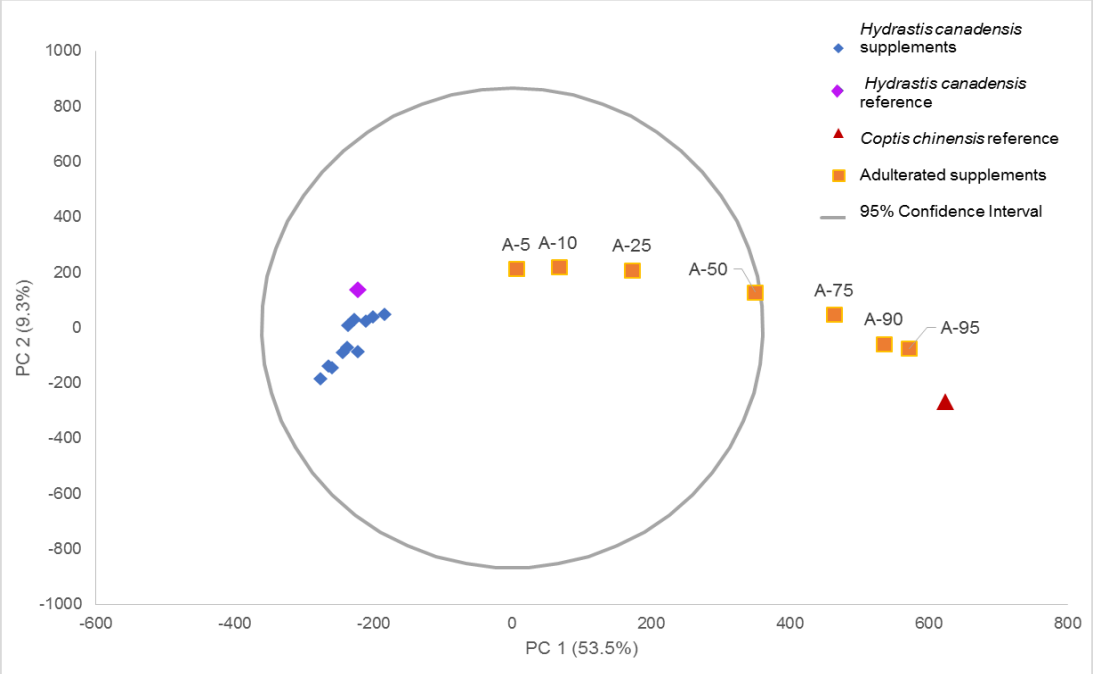
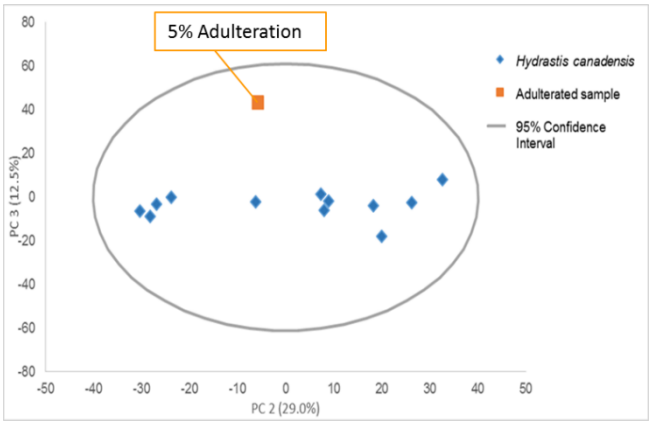
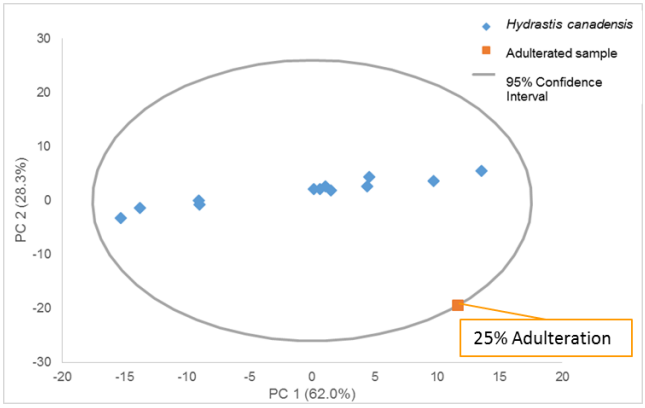


Figure S3. Highest Percentage of Adulterated Sample that was not an Outlier on all Platforms.

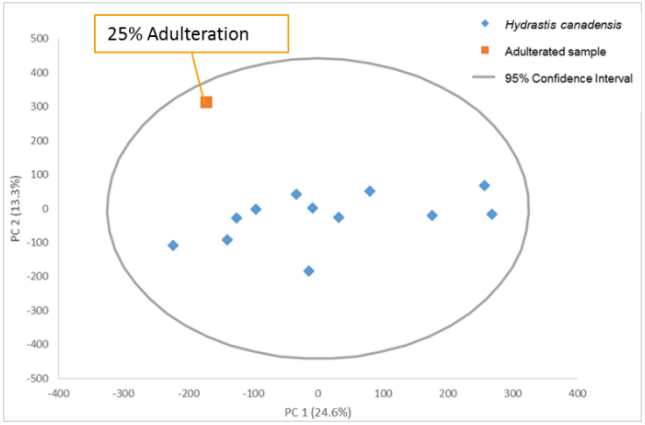
A



B

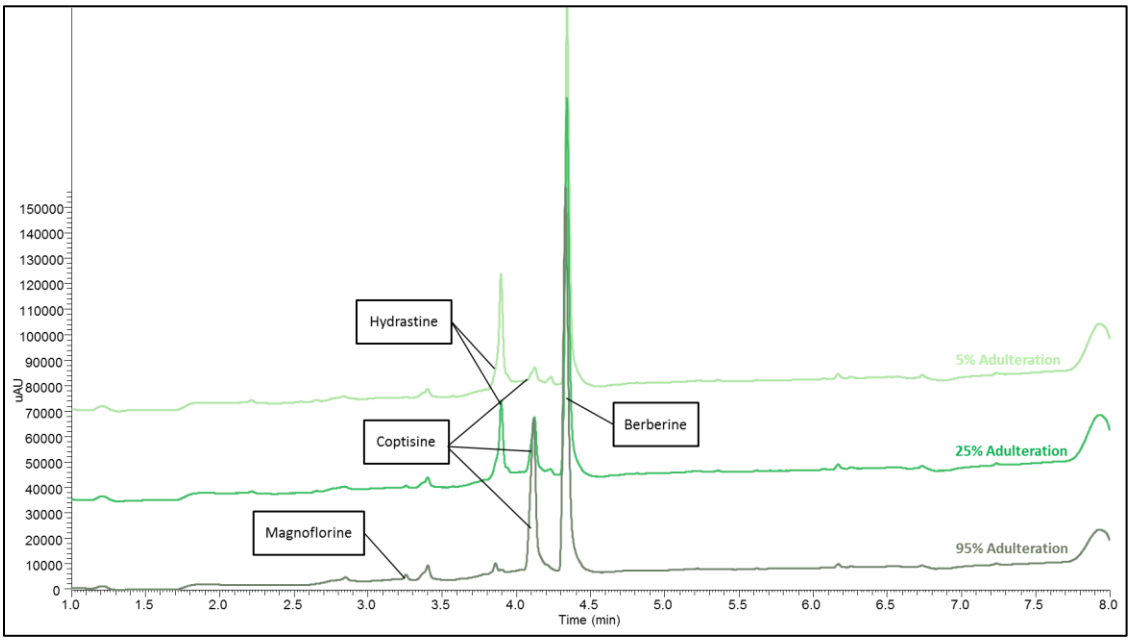


C



The lowest percentage not considered an outlier for the Orbitrap was 5% (A), for LC-UV 25% (B), and for the Q-ToF 25% (C).

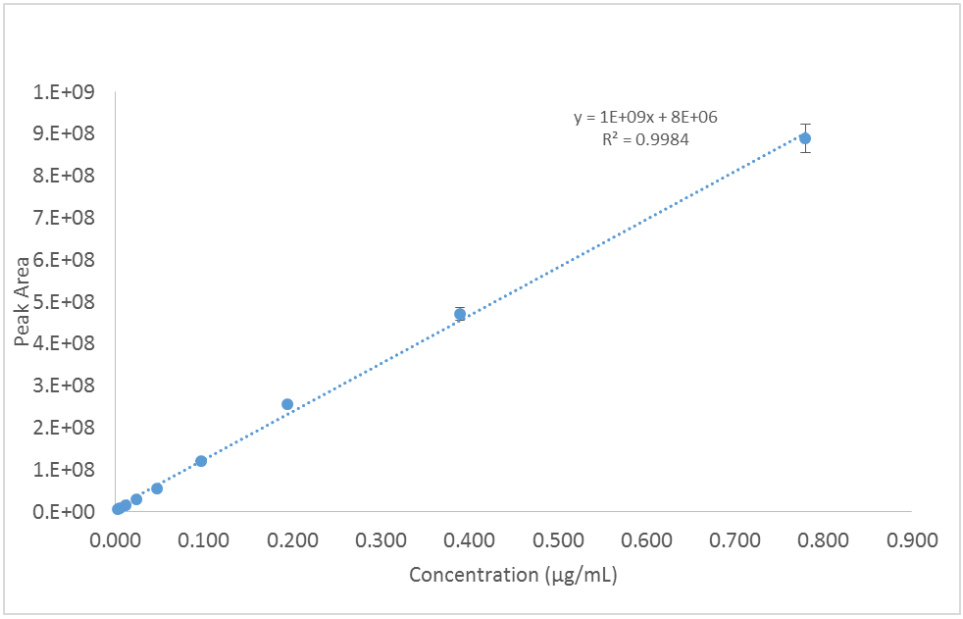
Figure S4. Stacked LC-UV Total Absorbance Chromatograms.



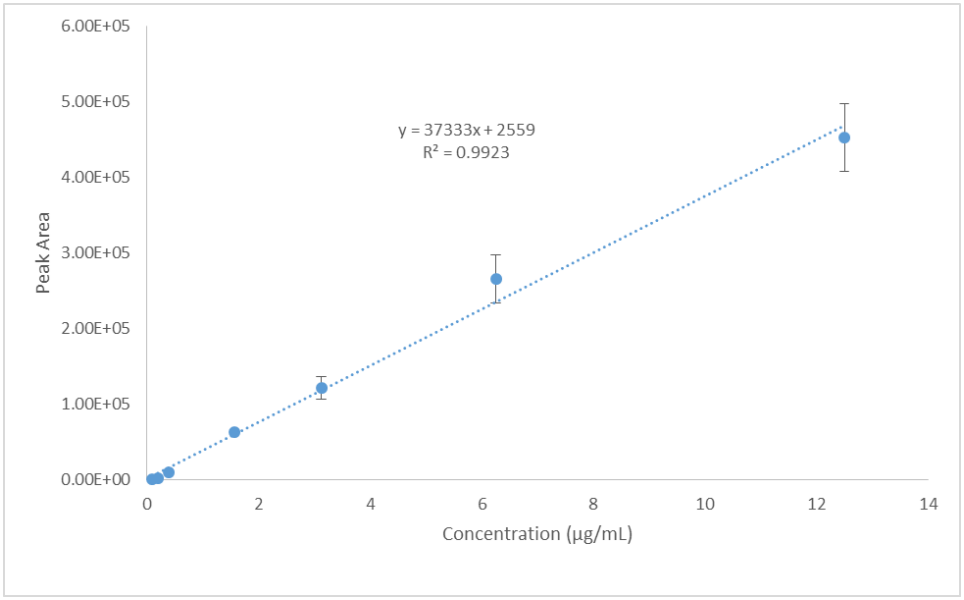
Though absorbance was not able to be separated and used as a variable, the total absorbance chromatogram (TAC) of each sample was used in the metabolomics. This stack of chromatograms shows the time versus the intensity of 5% adulteration (light green), 25% adulteration (green), and 95% adulteration (dark green). Peaks were labelled due to the retention time, absorbance, and *m/z* value obtained from the concurrent mass spectrometry analysis.

Figure S5. Calibration Curves for Palmatine.

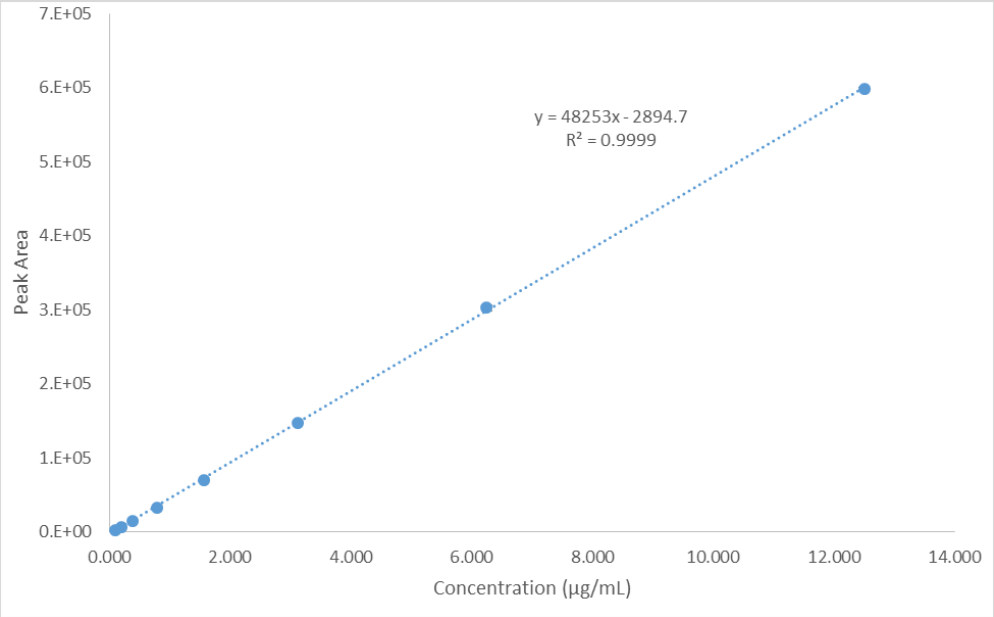
A



B

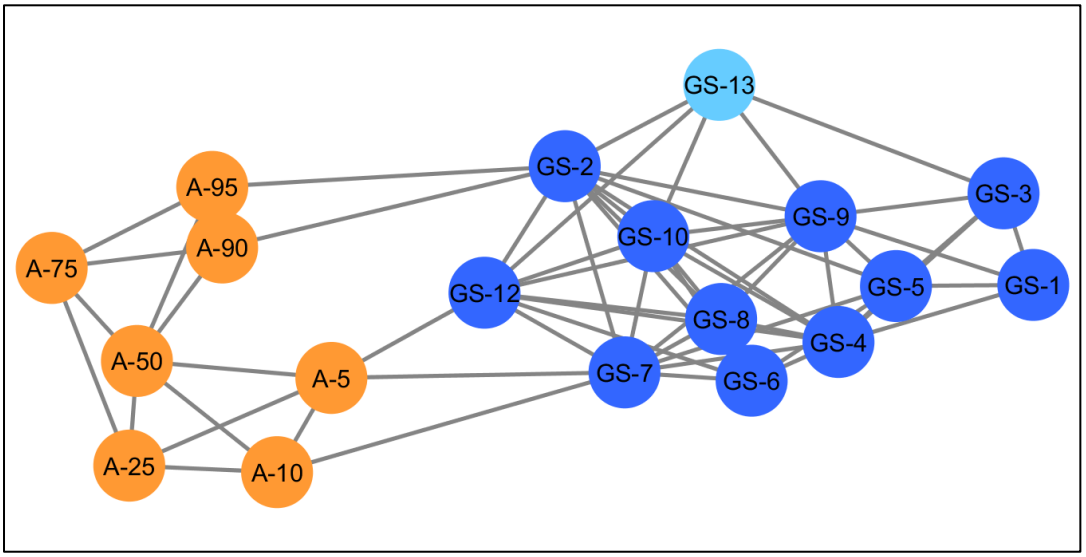


C



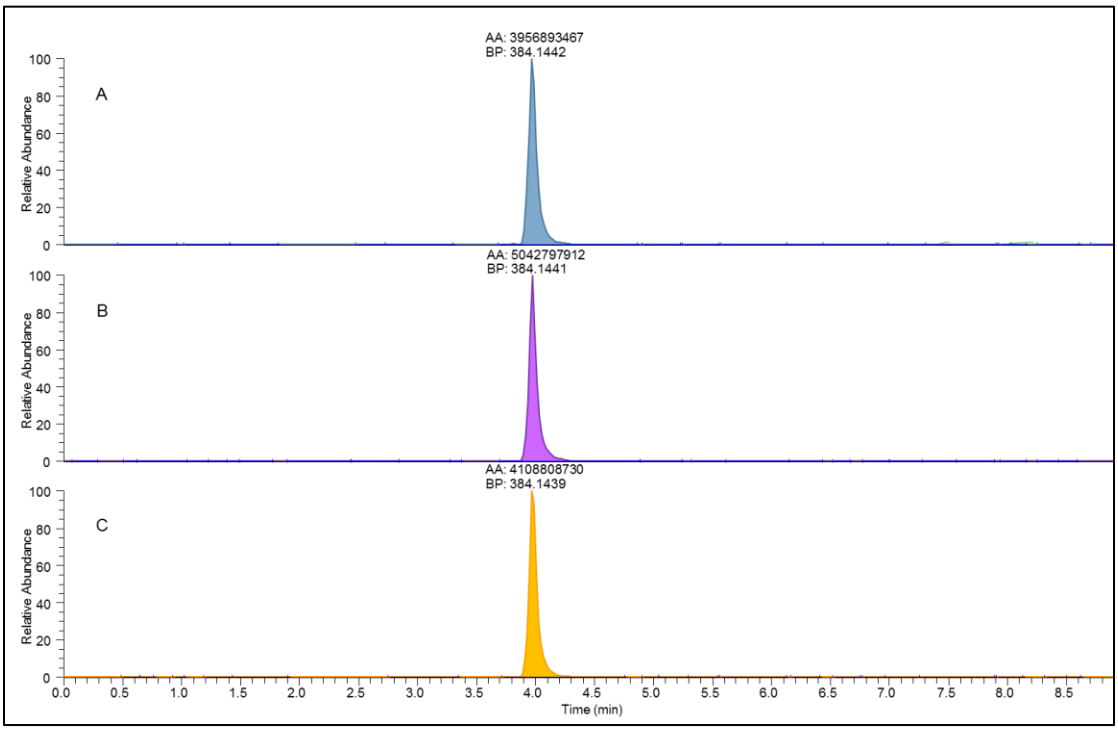
The Orbitrap (A) calibration curve was executed using lower concentrations. The Q-ToF (B) and LC-UV (C) curves both use the same concentrations of palmatine. Error bars are indicative of the standard deviation of technical replicates of each concentration.

Figure S6. Composite Score Analysis Plot with all Positive Connections.



This plot shows all positive connections with a similarity score above 0. This shows that there are still connections within the two groups, which is feasible given *H. canadensis* and *C. chinensis* are from the same family and have similar phytochemical profiles. However, it is clear that the two groups are still distinct at this level.

Figure S7. Peak Integration of Hydrastine.



Each peak corresponds to hydrastine, all of these are extracted ion chromatograms. The outlier in the SIMCA-PCA (A) has a peak area closer to the 5% adulterated sample (C) than the *H. canadensis* reference material (B). This is causing the product to be an outlier in regards to the SIMCA model. The product is still within the 95% confidence interval but is positioned slightly above the RSD value cutoff showing it varies slightly from the rest of the supplement group.