CHEMICAL CHARACTERIZATION AND QUALITY CONTROL OF AFRICAN NIGHTSHADE VEGETABLES

By

BO YUAN

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Food Science

Written under the direction of

Qingli Wu

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

October 2020

ABSTRACT OF THE DISSERTATION

CHEMICAL CHARACTERIZATION AND QUALITY CONTROL OF AFRICAN NIGHTSHADE VEGETABLES

by

BO YUAN

Dissertation director:

Dr. Qingli Wu

African indigenous vegetables (AIVs) play important role in providing the needed food security, nutrition and economic opportunities in sub-Saharan Africa. African nightshades, including *Solanum scabrum*, *S. nigrum*, *S. americanum* and *S. villosum*, and others are among the most popular leafy green vegetables. Yet, seed companies and the vegetable industry have largely ignored and undervalued these indigenous leafy greens in favor of the more traditional European introduced vegetables. As a consequence, the nutritional factors of the African nightshades are not well understood and have been studied little compared to European centric vegetables. In addition, many *Solanum* species are known to contain toxic glycoalkaloids, and such concern is also associated with *Solanum* African nightshades. This dissertation is devoted to the study of the nutritional and anti-

nutritive factors in African nightshades to identify the most nutritious edible nightshades, provide guidance for safe and nutritious consumption, and to provide the needed scientific knowledge to facilitate the proper promotion of these vegetables in sub-Saharan Africa and extended regions in developing countries to reduce hunger and improve nutrition. Such outcomes could also foster the creation of new agricultural and economic opportunities in these areas.

To accomplish these objectives, we first collected a wide array of genetic materials of the edible nightshades and then analyzed and chemically profiled them using various liquid chromatography coupled with mass spectrometry (LC-MS) methods and associated techniques that were developed over the course of this research. The methods and techniques developed were efficiently applied for investigation and quality control of the phytochemistry in the leaves and berries of African nightshades from different genetic sources and cultivation environments.

Chapters II to V are devoted to the chemistry of African nightshade leaves. In chapter II, the major phytochemicals were identified in the leaves, including phenolic acids such as chlorogenic acid; 23 glycosides of quercetin, kaempferol and rhamnetin; eight saponins of diosgenin, tigogenin and other analogues; and two glycoalkaloids of solasodine. In Chapter III, a phytochemical quantification method was developed and validated, which applied optimized acid-assisted hydrolysis to release the aglycones which were then quantified in tandem mass spectrometry (MS/MS). The impact of genetic sources and cultivation environment on phytochemical profile was also investigated and discussed. The results from all samples investigated showed that the leaves were safe for consumption due to the absence or very low content of glycoalkaloids and other anti-nutritives. Chapter IV

focuses on the free amino acids in the edible leaves. A hydrophilic interaction (HILIC)-LC-MS/MS method was developed and validated for determination of free amino acids in African nightshades as well as other AIVs including Ethiopian mustard, spider plant and amaranth. Different machine learning methods were employed for AIV classification prediction based on the profile of free amino acids. An online dashboard was also constructed for interactive application. Chapter V examines the vitamin-A precursor compound beta-carotene as well as other micronutrients to gain further insight into the overall nutritional contribution of edible nightshades.

Chapters VI to VIII shifted the research focus from leaves to berries of the African nightshades. While Africans normally consume only the leaves in edible nightshades, other groups such as those in South American value and consume the berries and not the leaves, despite the recognition of the presence of glycoalkaloids. Chapter VI was dedicated to phytochemical identification. Here, a total of 54 phytochemicals were identified, including phenolic acids of chlorogenic acid and neochlorogenic acid; flavonol glycosides of quercetin and isorhamnetin; anthocyanins of petunidin, malvidin and delphinidin; and saponins of diosgenin and tigogenin. In particular, a range of glycoalkaloids of solasodine and its uncommon and potentially novel hydroxylated and methylated derivatives were discovered, with the structure putatively identified based on the structural scaffold-fragmentation pattern. Chapter VII focused on quantity determination or estimation of the identified compounds in differently sourced berries, and discussed the profile change across different berry maturation stages. While many genetic lines were found to possess toxic levels of glycoalkaloids, a few genetic lines were found to be lacking in such toxins and were found instead to be rich in polyphenols. Such lines may be promising as new

foods in sub-Saharan Africa. As additional effort to enhance berry inspection and quality control, in Chapter VIII, a specialized novel in-source fragmentation MS/MS method was developed. This new high-throughput and sensitive method could be readily applied to rapidly distinguish safe from toxic berries.

Acknowledgement

I would like to express my special regards to my dissertation advisors Professors Qingli Wu and James E. Simon for their valuable guidance and remarkable support through my PhD study and dissertation, which has been an extraordinary journey in scientific discovery full of passion and dream in the wildest and most imaginary manner. None of my work would be ever possible without my advisors' diligent devotion to the lab and to my work. Working in this lab has been the single most important and fortunate decision I ever made during my stay in Rutgers University.

I thank Professors Chi-Tang Ho and Ilya Raskin for serving on my dissertation committee and for their professional critique on my work. Thanks are given to Dr. Fekadu Dinssa, WorldVeg Center, Arusha, Tanzania, plant breeder working with Prof. Simon on the improvement of edible nightshades for his leadership on the overarching project on African indigenous vegetables, and Professor Nicholas T. Bello for providing guidance in other associated research projects.

I wish to thank Dr. Danyue (Daisy) Zhao for being a dear colleague and friend of mine during her three-year postdoctoral training in this lab. It's been a great journey working with her through various interesting and challenging projects. Best wishes in her work in Hong Kong.

My dissertation benefited much from the insights and assistance from my lab mates. In particular, David Byrnes and William Reichert provided generous help with the field trail. Thomas Villani, Ariane Vasilatis and Daniel Giurleo developed the technique for carotene and vitamin E analysis, which was applied for nightshades analysis in Chapter V. Weiting Lyu co-constructed with me the content of Chapter VIII. Dushyant Kshatriya,

Daisy and I worked closely on the raspberry ketone project, part of the collaborative fruits presented in the Appendix B to E.

I whole-heartedly and respectfully appreciate the scholarship donated by Professor Henryk Daun from Food Science Department. I thank the scholarship presented by New Jersey Association for Food Protection.

I am indebted to the unconditional love and relentless support from my parents Xianxiu Liu and Peng Yuan. They're hard working doctors and dearest parents. I am grateful for my wife Ruoyuan Du for her company and support through the many difficult years. I love you all.

Last but not the least, I would like to acknowledge that part or whole portion of the following chapters have been published by time that I completed my dissertation in order of: Chapters II and V, in *Journal of Food and Drug Analysis (JFDA)*, 2018 [1]; Appendix B and C, in *JFDA*, 2018 [2]; Chapter VI, in *Journal of Food Science* , 2019 [3]; Chapter VII, in *Journal of the Science of Food and Agriculture*, 2019 [4]; and Chapter III, in *Journal of Food Chemistry (JFC)*, 2020[5]. In addition, Chapter VIII has been submitted to *JFC* and Appendix D and E submitted to Journal of Chromatography B currently under peer review.

Published Chapters

[1] **Yuan B**, Byrnes D, Giurleo D, Villani T, Simon JE, Wu Q. Rapid screening of toxic glycoalkaloids and micronutrients in edible nightshades (*Solanum* spp.). J Food Drug Anal 2018;26:751-760. doi.org/10.1016/j.jfda.2017.10.005

[2] **Yuan B**, Zhao D, Du R, Kshatriya D, Bello NT, Simon JE, Wu Q. A highly sensitive ultra-high performance liquid chromatography/tandem mass spectrometry method with in-source fragmentation for rapid quantification of raspberry ketone. Journal of Food and Drug Analysis, 27(3), 778-785. https://doi.org/10.1016/j.jfda.2018.07.005

[3] **Yuan B**, Byrnes D, Dinssa FF, Simon JE, Wu Q. Identification of Polyphenols, Glycoalkaloids, and Saponins in *Solanum scabrum* Berries Using HPLC-UV/Vis-MS. J Food Sci 2019;84:235-243. doi: 10.1111/1750-3841.14424..

[4] **Yuan B**, Byrnes D, Dinssa FF, Simon JE, Wu Q. Quantity assessment of polyphenols, glycoalkaloids and saponins in *Solanum scabrum* berries of different genetic sources and maturity by HPLC/UV-visible/MS methods. J Sci Food Agric 2019;99:3578-3587. doi: 10.1002/jsfa.9579.

[5] **Yuan B**, Dinssa FF, Simon JE, Wu Q. Simultaneous quantification of polyphenols, glycoalkaloids and saponins in African nightshade leaves using ultra-high performance liquid chromatography tandem mass spectrometry with acid assisted hydrolysis and multivariate analysis. Food Chem 2020; 312:126030. doi: 10.1016/j.foodchem.2019.126030

Table of Contents

List of Illustrations

List of Tables

List of Appendix Illustrations

List of Appendix Tables

CHAPTER I. INTRODUCTION

1 Background

1.1 African indigenous vegetables

African indigenous vegetables (AIVs), also called traditional African vegetables, play important roles in food security, nutrition, food diversity, rural and economic development and sustainable land care [1]. Such AIVs, collected from the wild or cultivated to a limited extent provide vital food and nutrient sources and also important income generating opportunities for the typical small-scale farmers, especially in economically limited regions [2]. Many AIVs are not only consumed but also utilized for their preventive and/or therapeutic medicinal effects. Adapted to the local environment, AIVs often provide more sustainable production than exotic or introduced crops such as European vegetables [3]. Some of the most important AIVs include nightshades (such as *Solanum scabrum* and *S. nigrum* etc.), amaranth (*Amaranthus cruentus*), eggplant (*Solanum aethiopicum*), okra (*Abelmoschus caillei*) and jute mallow (*Corchorus olitorius*). Efforts are being made to increase the farming and marketing of AIVs in an attempt to alleviate hunger and improve nutrition, and to increase farmers' income, improving the local and regional economy [3].

Many of nightshade species are considered agricultural weeds in North America and Europe, and the ruptured berries with its deep purple color are also treated as a source of contaminants if inadvertently harvested with other crops [4]. In contrast, many of the African nightshade species, are among the most popular and preferred leafy AIVs in the sub-Saharan area, and their leaves, as well as tender stems and young shoots, are prepared and consumed like spinach and amaranth.

The African nightshades represent a wide group of botanically and genetically related plants and constitute approximately 30 species, collectively belonging to the *Solanum* section, also known as section *Morella*, which belongs to the larger *Solanum* genus in the *Solanaceae* family [5, 6]. These nightshades are diversely referred to as vegetable nightshades, edible nightshades, garden nightshades, common nightshades, '*S. nigrum* complex', or '*S. nigrum* and related species [5, 7]. Species belonging to the *Solanum* have a wide distributed, from sea level to altitudes over 3500 meters, and from temperate to tropical areas [5]. In Africa, some of the most economically important nightshades include *S. scabrum*, *S. americanum*, *S. villosum* and *S. nigrum*.

Solanum scabrum is one of the most important edible nightshades in sub-Saharan Africa, with the highest leaf, berry and seed yields [1] among African nightshades. It is synonymous with *S. melanocerasum*. All, *S. guineense* (L.) Miller and *S. intrusum* Soria, and is sometimes referred to by the common name "garden huckleberry", which though having the same common name is distinguished from the horticultural huckleberry (*Vaccinium* spp.) [5, 8]. *S. scabrum* features prominently and dentately winged stems, large ovate leaves with sinuate margins, stellate white or tinged purple flowers, and broadly ovoid and enriched purple fruits that remain on plant at maturity [5].

The leaves of *S. scabrum* are an important source of vegetables consumed and/or marketed in sub-Saharan area. The berries, despite the prolific production, however, are generally discarded by Africans, possibly because of a perception of toxicity due to bitterness and the associated toxic glycoalkaloids in the berries [6, 9]. This is in stark contrast to practices in regions of South America and Europe, where the berries are consumed fresh or made into juices, sauces and jams or colorants [10, 11] and the leaves

discarded. This stark consumption differences between different geographical regions and cultural practices may result from berry variation in toxicity and nutrition due to genetic and/or environmental differences, because of cultural factors or misconception of berry edibility, or because historically the South Americans and/or Europeans discovered postharvest techniques to remove or reduce the alkaloid content during food preparation. The insufficiency of studies and the lack of clarity as to this plant safety while popularly consumed could lead to acute or long-term health hazard; or if properly processed or through plant breeding made to have trace or no alkaloids, then the berries are an untapped food supply in sub-Saharan area.

1.2 Phytochemistry

In this section, the most common types of phytochemicals found in the plant kingdom are briefly reviewed, followed by a more specific discussion of the phytochemicals found in African nightshades.

1.2.1 Polyphenols

Polyphenols are a broad spectra of plant secondary metabolites, bearing an aromatic ring with one or more hydroxyl groups on it, with the structure ranging from simple phenolic molecules to highly complicated polymers. This group of natural products that can be roughly divided into flavonoids, phenolic acids, stilbenes, tannins and lignans.

1.2.1.1 Flavonoids

Flavonoids typically assume a C6-C3-C6 configuration, with two aromatic rings (A and B) linked by a C-3 bridge. Based on variation of the C ring structure, i.e., existence or not of the C2-C3 double bond, 4-C carbonyl group and 3-hydroxyl group, as well as other special features, flavonoids can be categorized into flavone, flavonol, flavanone, flavan-3-ol, flavanonol, isoflavone, chalcone, and anthocyanidin, as shown in **Figure I-1**.

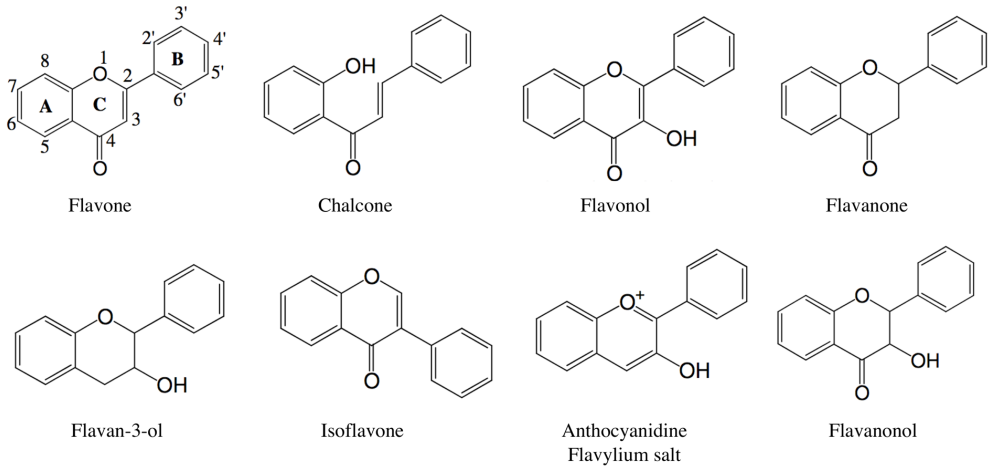


Figure I-1. Representative structures of flavonoids in plants (adapted from [13])

1.2.1.2 Phenolic acids

Phenolic acids account for about one-third of dietary polyphenols, and could exist in free forms or bound to various components via ester, acetal and ether bond [12]. Phenolic acids can be divided into derivatives of benzoic acids and those of cinnamic acids. Some of the most common benzoic acid based phenolic acids include gallic acid, hydroxybenzoic acid, protocatechuic acid, vanillic acid and syringic acid, typicall bearing a C6-C1 skeleton.

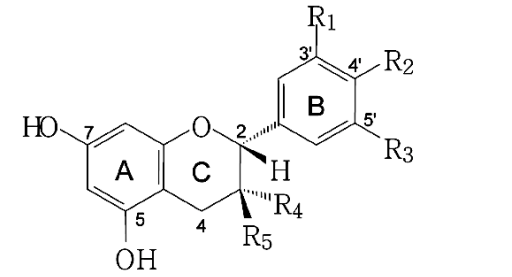
Some of the common cinnamic acid-based phenolic acids include caffeic acid, ferulic acid, coumaric acid and sinapic acid, typically assuming a C6-C3 structure.

1.2.1.3 Tannins

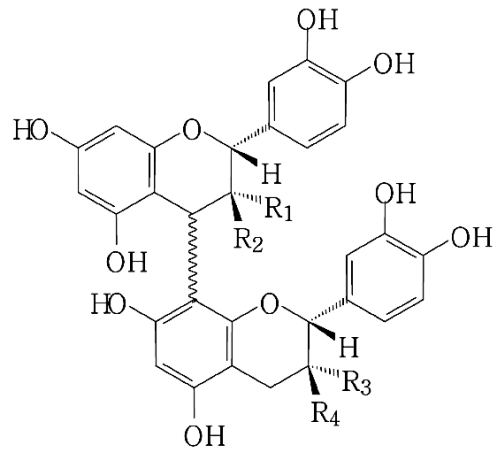
Tannins can be further divided into unhydrolyzable tannins and hydrolysable tannins. Unhydrolyzable tannins are also common referred to as proanthocyanidins. By formation of complexes with salivary proteins, these compounds are responsible for the astringency character of fruits. Proanthocyanidins are oligomers of polymerized flavan-3-ols, formed most commonly by C4-C8 linkage as well as C4-C6 linkage, both called the type B structure. If additional ester bond is formed between the C2-O-C7, then the linkage is referred to as type A. Typical structures are presented in **Figure I-2**. Proanthocyanidins exclusively composed of (epi)catechin are referred to as procyanidin (same substitution pattern in the A ring and B ring as the *anthocyanidin* cyanidin) [14].

Hydrolysable tannins are derivatives of gallic acids that are bond to one or more polyols, and the galloyl groups can be further esterified or crosslinked to form more complex structure. Depending on the crosslink pattern of the gallic acids involved, hydrolysable tannins may be further classed into gallotannins and ellagitannins. In gallotannins, gallic acids are usually esterified to a polyol, where glucose being the most common; one gallic acid may be also esterified with another gallic acid via formation of the meta- or para- depside bond. Representative structures are shown in **Figure I-3A**. In ellagitannins, however, two gallic acids are crosslinked via C-C bond to form hexahydroxydiphenic acid (HHDP) as the repeating basis, and HHDP continues to form esters with other polyol units. Upon hydrolysis and HHDP is released from the polyol

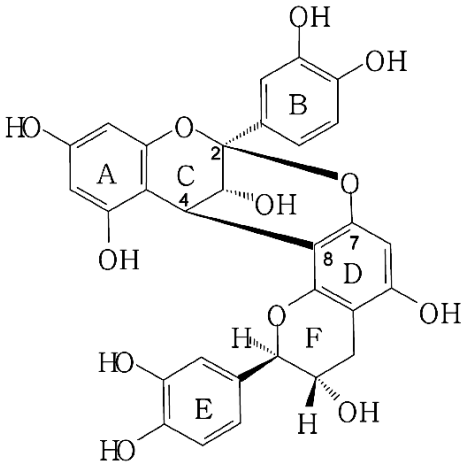
groups, HHDP spontaneously lactonize into ellagic acids. Typical structures are shown in **Figure I-3B**. [15]



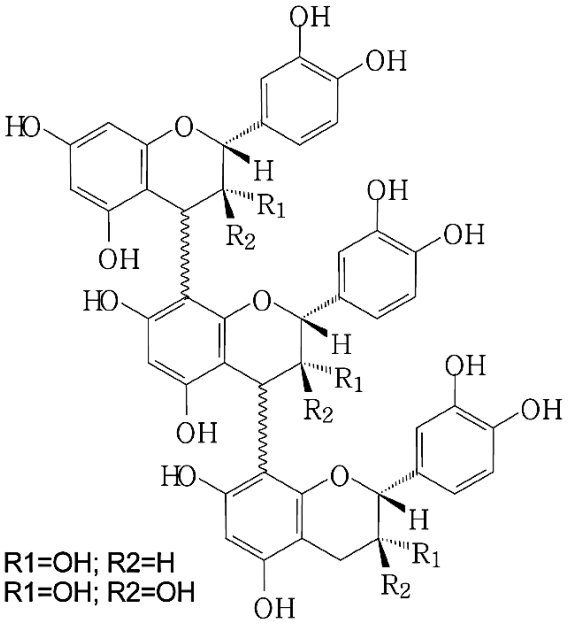
Flavan-3-ols:	R_1	R_2	R_3	R_4	R_5
Afzelechin:	H	OH	H	H	OH
Epiafzelechin:	H	OH	H	OH	H
Catechin:	H	OH	OH	H	OH
Epicatechin:	H	OH	OH	OH	H



B1: R1=OH; R2=H, R3=H, R4=OH
B2: R1=OH; R2=H, R3=OH, R4=H



Dimer A2.



C1: R1=OH; R2=H
C2: R1=OH; R2=OH

Figure I-2. Representative structure of proanthocyanidins found in nature. Structures of the monomeric flavan-3-ols, dimer B1 and B2, trimer C1 and C2, and dimer A2 are shown (cited from [14]).

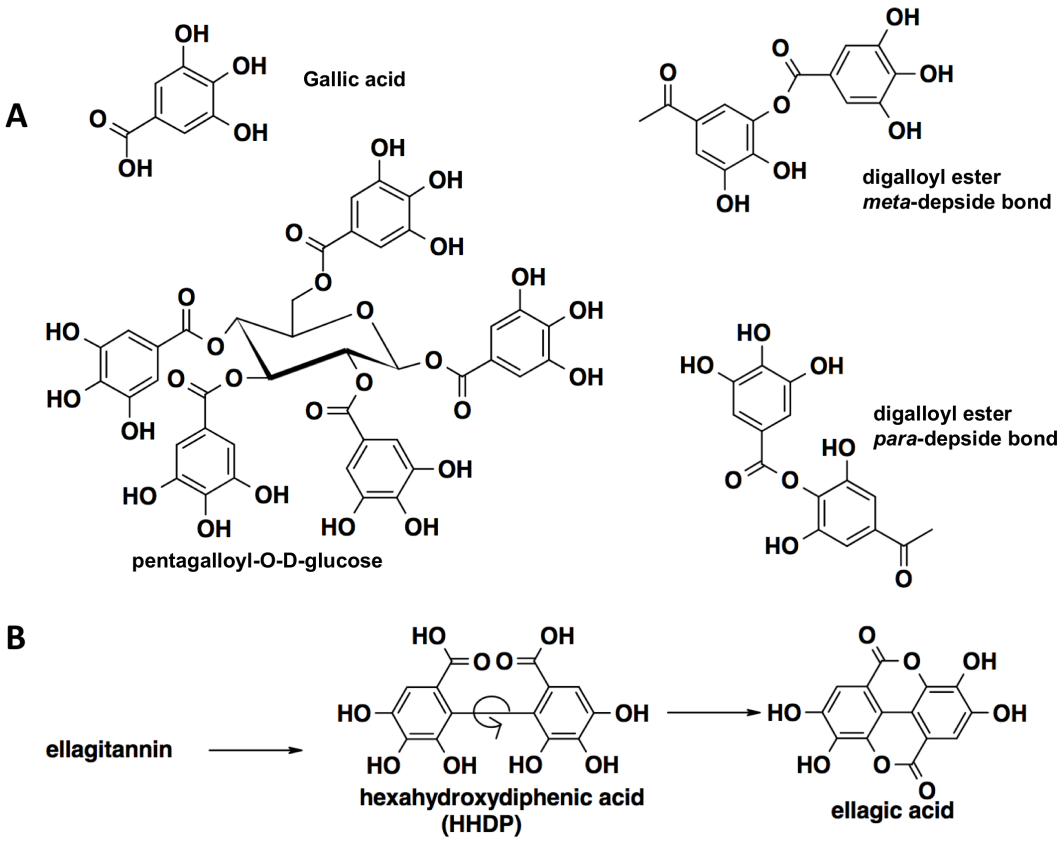


Figure I-3. Representative structures of hydrolysable tannins in plants. (**A**), structure of gallotannins. (**B**), basic repeating unit of hexahydroxydiphenic acid (HHDP) in ellagitannins and generation of ellagic acid upon hydrolysis (adapted from [15]).

1.2.1.4 Polyphenols identified in African nightshades

Elaborated and systematic work on nightshade polyphenols (especially for the leafy parts) are scares in literature; if any, many of these work largely applied chemical-assay & color-changing based screening methods, and the preliminary nature did not allow for deeper insight into polyphenol constitution [16, 17].

In one pioneering study by Huang et al using HPLC/UV by comparing with the authentic standards, a wide range of polyphenols were detected in leaves, stems, raw and green fruit and mature purple fruit of *S. nigrum* [18]. The polyphenols "identified" in this

work includes phenolic acids of benzoic acid-derived phenolic acids, i.e., gallic acid, gentisic acid, protocatechuic acid, vanillic acid, syringic acid; cinnamic acid-derived phenolic acids, i.e., caffeic acid, *p*- and *m*-coumaric acid, ferulic acid, and chlorogenic acid; flavan-3-ols and derivatives, i.e., catechin, epicatechin, epigallocatechin gallate, gallocatechin gallate; flavonols, i.e., rutin; flavanone, i.e., narigenin and hesperetin; flavone, i.e., luteolin and apigenin; flavonol, i.e., myricetin, quercetin and kaempferol; and anthocyanidins of petunidins, delphinidins, pelargonidin, and trace level of peonidin and malvidin, detected after hydrolysis. Pioneering and valuable as this work is, compound identification by mere usage of retention time comparison with reference standards as conducted in this work (especially for leafy and stem part) might remain ambiguous, given the lack of specificity/selectivity of UV 295 nm used for non-anthocyanidin polyphenols, presence of high background noise and interference peaks and associated less ideal chromatographic resolution, and low intensity for some of the peaks of interest. In addition, reporting the aglycones *per s*e yet without and/or lacking detection of the corresponding glycosides seems somewhat unsatisfactory, considering that most flavonoids detected exist in the form of glycosides more than free aglycones.

Compared with the leafy parts of nightshades, berries attracted much more intellectual interest relative to polyphenol study [10, 19]. In the most recent work using UPLC- Q/TOF-MS on berries of *S. scabrum*, anthocyanins of malvidin, petunidin, and delphinidin with similar glycosylation and acylation (with ferulic and coumaric acid) pattern were identified, with petunidin-3-(*p*-coumaroyl-rutinoside)-5-O-glucoside being the single most important anthocyanin [10], which was likely petanin, the same major anthocyanin in blue/purple potatoes [20]. Apart from anthocyanins, two other flavonoids species, i.e., quercetin mono-

glycosylated compounds were detected in relatively low amount in the *S. scabrum* berries. Esters of quinic acid with caffeic acid (i.e., the (neo)(crypto) chlorogenic acids) or with coumaric acid were also detected; some of them were further acylated with acetic or malonic acid [10].

1.2.2 Alkaloids

1.2.2.1 Chemistry of Solanum Alkaloids

Many *solanum* species are known to contain toxic glycoalkaloids. Many of the glycoalkaloids are shared in common across multiple species, and could also be unique or limited to a few species. The glycoalkaloid, with its amphiphilic nature, are composed of two units, one unit being the hydrophobic 27-carbon skeletal aglycone with the nitrogen atom in the F ring, the other being the hydrophilic carbohydrate side chain attached on the 3-OH position. The aglycone, also called the alkamine, could be structurally divided into five categories: solanidanes, with the fused indolizidine structure composed of the E-F rings; spirosolanes, featuring an oxa-aza spiro structure; and (22,26)-epiminocholestanes, α-epiminocyclohemiketals, and 3-aminospirostanes, as presented in **Figure I-4**. The first two, soladidane and spirosolanes, are the most common skeleton structures in *Solanum* [21].

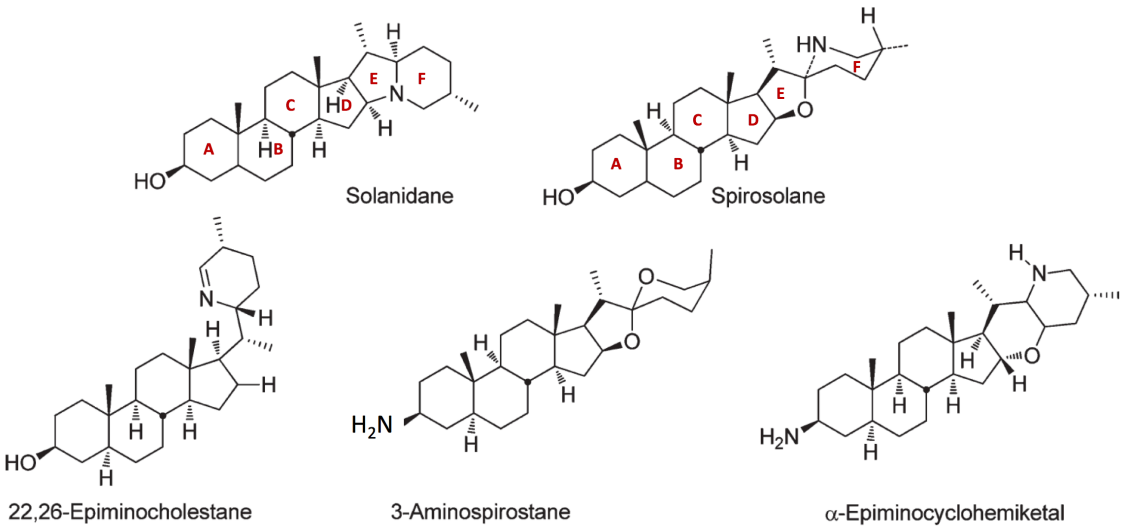


Figure I-4. Representative skeletons of steroidal alkaloids in *Solanum* spp. (cited and adapted from [21]). Typical ring notations are marked in the solanidane and spirosolane structures.

Some of the common and typical structures of glycoalkaloids are presented in **Figure I-5**, including solasodine-based glycosides, such as α-solamargine and α-solasonine, the major glycoalkaloids in *S. melongena* or eggplants; solanidine-based glycosides, such as α-chaconine and α-solanine, the principal glycoalkaloids in *S. tuberosum* or potatoes; and tomatidine-based glycosides, such as tomatine as found in S. *lycopersicum* or tomatoes. Acid-assisted hydrolysis of these glycoalkaloids renders products of the mono-, di-, and trisaccharide derivatives and the corresponding aglycones.

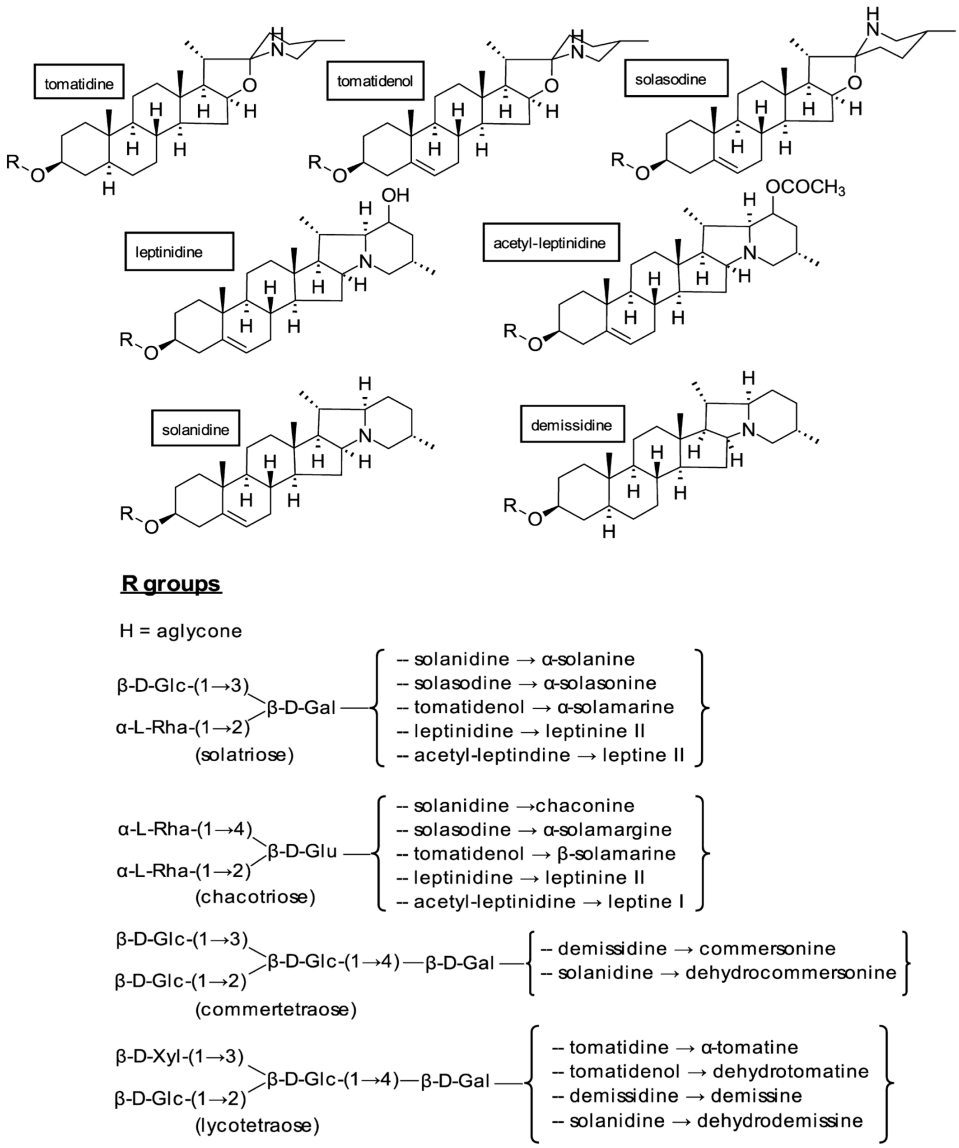


Figure I-5. Typical structures of glycoalkaloids in *Solanum* spp. (cited from [34]). For saccharide abbreviations, Glc, glucose; Rha, rhamnose; Xyl, xylose; Gal, galactose.

1.2.2.2 Alkaloids identified in nightshades

Solanum nigrum was extensively reported as a source of glycoalkaloids. The total content of alkaloid was reported in one study to be ~ 1.6 mg/100 g dry mass in leaves and ~1.1 mg/100 g in the seeds, yet the identity of the alkaloids quantified remained unclear due to limit of methods applied [17, 22]. In another study, solasodine glycosides of varied carbohydrate side chain (each 2 ~ 180 mg) and a less common 12-OH solasodine glycosides (~ 38 mg) were extracted from the aerial parts (10 kg dry mass) and identified [23]. And solasodine glycosides continued to be detected as major glycoalkaloids in *S. nigrum* aerial parts by other researchers [24]. Apart from solasodine compounds, other derivatives were occasionally identified using NMR: 27-dihydroxysolasodine (C and F ring substitution) was isolated either as a free aglycone or glycosides from the berries [25, 26], and *N*-methylsolasodine though very rare was also reported [27].

For *S. scabrum*, solamargine and solasonine arguably the most important alkaloids as in *S. nigrum* was extracted and separated from both leaves and fruits, with the fruits being more enriched with the glycoalkaloids; solanidine and tomatidenol, the two aglycones *per se*, were also detected after hydrolysis, suggesting potential existence of corresponding glycosides [28]. In another study, the existence of alkaloids in the leaves were decided to be “doubtful” [29]. In one recent report, screening by LC-QTOF-MS revealed a lack of glycoalkaloids in the methanol leafy extracts of both *S. scabrum* and *S. villosum* [30].

1.2.3 Saponins

Diosgenin and tigogenin, the oxygenous counterparts of solasodine have long been recognized as two of the major sapogenins in many nightshade species, with diosgenin being usually the most abundant [6]. Most of earlier work, as also in the case of glycoalkaloids, typically applied acid-assisted hydrolysis to yield the aglycones for subsequent structural elucidation and/or quantification (e.g., colorimetric reaction, GC and GCMS), with limited focus on the carbohydrate chain. More recent years have witnessed expanding number of studies applying direct isolation and purification of the individual glycosides for new structural discovery and/or for use as standards for accurate quantification [31-33]. One of the landmark studies by Wang et al. isolated a total of 17 saponin compounds (**Figure I-6**) from the unripe berries of *S. nigrum*, covering most of the typical aglycone structures commonly found in nightshades, though surprisingly, no diosgenin nor tigogenin based glycosides were detected [33].

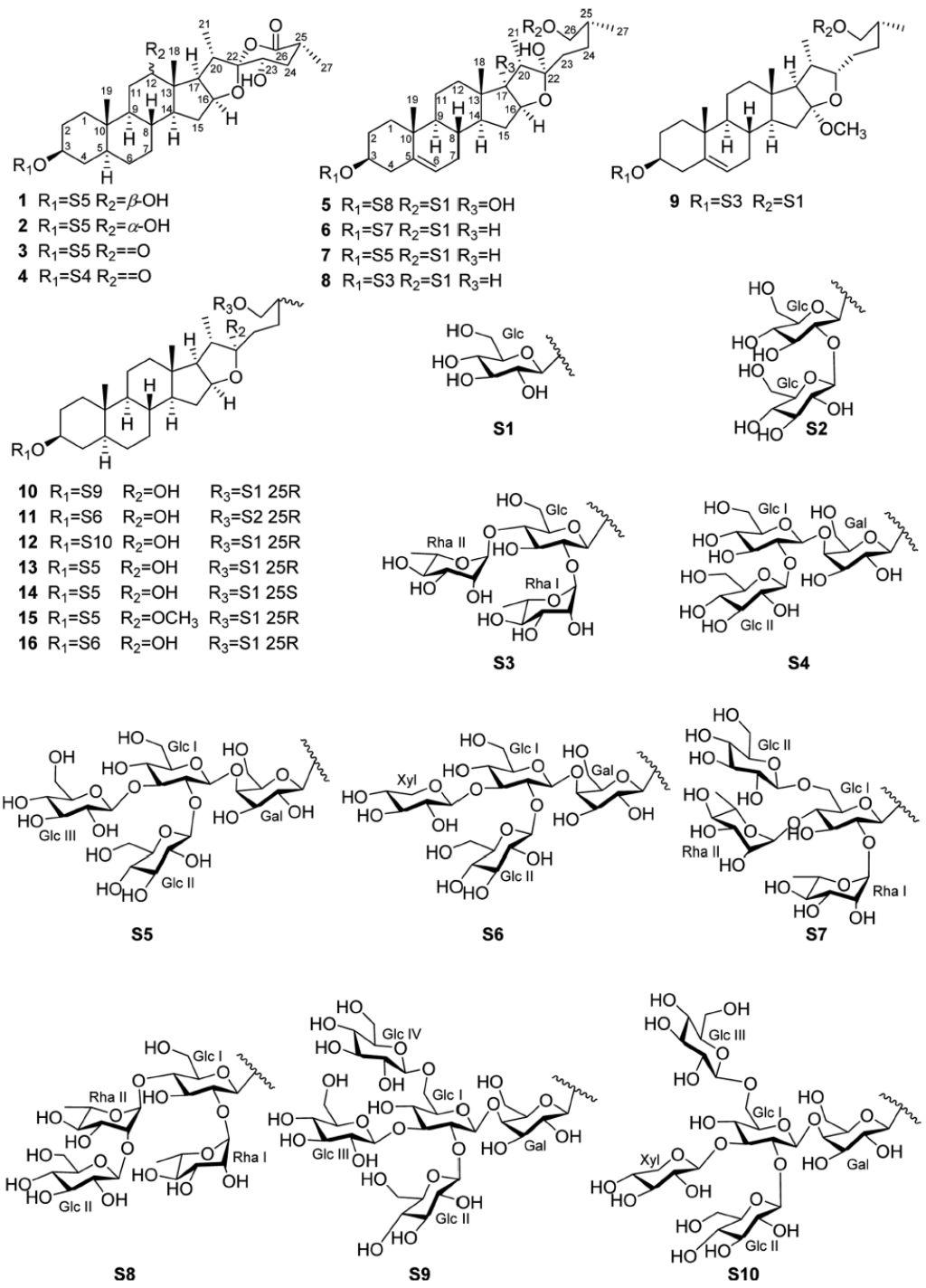


Figure I-6. Saponin structures isolated from unripe berries of *Solanum nigrum* (cited from [33]).

1.2.4 Protein

AIVs are a valuable source of vegetable proteins. In one study reported by Kamga et al. [35], the crude protein content, based on total nitrogen measurement, accounted for 9 %~ 38 % of dry mass in a total of five commonly consumed AIV species investigated, including amaranth (*Amaranthus cruentus*), eggplant (*S. aethiopicum*), jute mallow (*Corchorus olitorius*), okra (*Abelmoschus callei*) and one nightshade species *S. scabrum*.

The *S. scabrum* was a most protein enriched AIV, with the protein content accounting for up to 33~38% of dry mass, and 100 mg dry mass, if complete protein assumed, could contribute to 49.5% of the daily protein requirement of the pregnant and lactating mothers [35]. In another study, the crude protein content in nightshade *S. nigrum* was reported ~24.9 % of dry mass [17].

Despite the ideally high content of leafy protein, however, the protein quality of nightshade leaves so far remains largely unknown. This, lack of understanding compromises the full utilization of the protein value due to possibly ill-balanced amino acid profile and unbalanced diet pattern. This lack of knowledge also limits those in plant breeding and crop improvement seeking to improve the nutritional value of this underrecognized leafy green. A deeper insight with this regard would help increase protein utilization; formulation with crops of complementary protein profile, such as with amaranth known to contain enriched essential amino acid lysine lacking in many staple foods [36], could enhance utilization of nightshade protein and overall nutrient value of the diet.

2 Rationale and Hypothesis

African indigenous vegetables (AIVs) play important roles in food security, food diversity, nutrition, and economic development in sub-Saharan Africa [7]. Particularly, African nightshades, belonging to the *Solanum* genus, are among the most popular and top priority leafy AIVs. Scientifically wise promotion of cultivation, consumption and marketing of the African nightshades to wider regions is of practical significance to combat against hunger, increase nutrition and enhance income in African rural area.

Solanum species are known to contain toxic glycoalkaloids [21], and the associated safety concern is also associated with the edible African nightshades, both leaves and berries, thus potentially compromising their safe consumption and recognition. In fact, most berries are discarded without consumption, remaining a prolific yet neglected agricultural resource in sub-Saharan area. It is of practical importance to systematically study the associated compounds to distinguish safe species and/or genotypes to ensure consumption safety and to discover new food supply.

Determination of nightshade glycoalkaloids has mostly relied on tedious purification of individual compounds [23, 26, 37, 38]; powerful though this conventional strategy is, it does not allow for routine analysis, and did not provide the complete and quantitatively accurate view of the holistic glycoalkaloid profile. Besides, the content of glycoalkaloids in nightshade species are complicated by differences among species, genotypes and growth environment. Apart from glycoalkaloids, other important secondary metabolites in African nightshades including polyphenols and saponins that are of nutritional and/or toxicological effects have not been adequately studied.

Development of modern instrumentation has greatly advanced compound separation and identification. Particularly, application of liquid chromatography hyphenated with mass spectrometry (LC-MS) has allowed for efficient compound separation and structural elucidation based on known scaffold-fragmentation pattern [39], and greatly facilitated study of bioactive compounds in plants and food systems. In addition, machine learning-based techniques coupled with the high-throughput analysis tools have allowed for acquisition of deeper insight into analytical data.

Based on abovementioned rationale, we hypothesized that LC-MS methods can be developed and/or validated for analysis of African nightshade bioactive compounds/nutrients and antinutritive factors. The development of such analytical approaches will facilitate an understanding of nutrition and toxicity of nightshade leaves and berries; and that the instrumental and associated statistical and computational methods to be established will provide analytical solution for routine analysis and quality control of nightshades as vegetables and/or fruits to ensure consumption safety.

3 Specific Aims

3.1 Leaves

3.1.1 Identification of phytochemicals in leaves

1) Development of HPLC-ion trap MS method for separation and identification of major phytochemicals in the leaves of African nightshades.

2) Development of HPLC-ion trap MS/MS method for profiling of aglycones released after acid-assisted hydrolysis for additional aglycone structural elucidation and confirmation.

3.1.2 Quantification of phytochemicals in leaves

1) Development of a simultaneous hydrolysis method for flavonol glycosides, glycoalkaloids and saponins in the leaves.

2) Development of a UHPLC-triple quadrupole (QqQ) MS/MS method to conduct glycosides and aglycones quantification associated with the hydrolysis/recovery-optimization study.

3) Development and validation of (i) a second UHPLC-QqQ-MS/MS method for quantification of all major aglycones in the leaves using hydrolysis method developed; (ii) a third supplementary UHPLC-QqQ-MS/MS method for quantification of aglycones *not* hydrolyzed.

4) Contrast and compare the aglycone profile from plants of different species, accessions/genetic sources and cultivation environments.

5) Evaluation of the nutrition and toxicity associated with the phytochemical profile studied.

3.1.3 Leafy free amino acids determination

1) Development and validation of a high-sensitivity hydrophilic interaction (HILIC) UHPLC-MS/MS method of free amino acids and validation with random effects model and nested design structure.

2) Quantification of free amino acids in different species of African vegetables using the HILIC-UHPLC-MS/MS method developed.

2) Application of different machine learning techniques to "learn" to distinguish different species based on amino acid profile.

3.2 Berries

3.2.1 Identification of phytochemicals in berries

1) Development of HPLC-ion trap MS methods to separate and identify major phytochemicals in the berries of African nightshades.

2) Development of HPLC-ion trap MS/MS method to separate and identify aglycones released after acid-assisted hydrolysis for additional aglycone structural elucidation and confirmation.

3.2.2 Quantification of phytochemicals in berries

1) Quantity assessment of berry phytochemicals by the method developed in prior section using representative reference standards of each category of compounds.

2) Compare and contrast the phytochemical profile in berries from different genetic sources and maturation stages.

3) Evaluation of the nutrition and toxicity associated with the phytochemicals investigated.

3.2.3 Rapid quality control for glycoalkaloids

Development of a fragmentation pathway-based UHPLC-QQQ-MS/MS with in-source fragmentation method for high-throughput, sensitive detection of glycoalkaloids in the berries for rapid quality control.

4 References

[1] Yang R-Y, Ojiewo C. African nightshades and African eggplants: taxonomy, crop management, utilization and phytonutrients. Afr Nat Plant Prod 2013;2:137-165.

[2] Weinberger K, Msuya JM. Indigenous vegetables in Tanzania: significance and prospects: World Vegetable Center; 2004.

[3] Mal B. Neglected and underutilized crop genetic resources for sustainable agriculture. Indian J Plant Genet Resour 2007;20:1-14.

[4] Ogg Jr A, Rogers B, Schilling E. Characterization of black nightshade (*Solanum nigrum*) and related species in the United States. Weed Sci 1981:27-32.

[5] Edmonds JM, Chweya JA. Black nightshades: *Solanum nigrum* L. and related species: Bioversity International; 1997.

[6] Carle R. Investigations on the content of steroidal alkaloids and sapogenins within *Solanum* sect. *Solanum* (= sect. *Morella*)(Solanaceae). Plant Syst Evol 1981;138:61-71.

[7] Yang R-Y, Ojiewo C. African nightshades and African eggplants: taxonomy, crop management, utilization and phytonutrients. Afr Nat Plant Prod 2013;2:137-165.

[8] Defelice MS. The black nightshades, *Solanum nigrum* L. et al.—poison, poultice, and pie. Weed Technol 2003;17:421-427.

[9] Gbile Z, Adesina S. Nigerian *Solanum* species of economic importance. Annals of the Missouri Botanical Garden 1988:862-865.

[10] Oszmianski J, Kolniak-Ostek J, Wojdylo A. Characterization of phenolic compounds and antioxidant activity of *Solanum scabrum* and *Solanum burbankii* berries. J Agric Food Chem 2014;62:1512-9.

[11] Francis F, Harborne J. Anthocyanins of the garden huckleberry, *Solanum guineese*. J Food Sci 1966;31:524-528.

[12] Ignat I, Volf I, Popa VI. A critical review of methods for characterisation of polyphenolic compounds in fruits and vegetables. Food Chem 2011;126:1821-35.

[13] Tapas AR, Sakarkar D, Kakde R. Flavonoids as nutraceuticals: a review. Tropical Journal of Pharmaceutical Research 2008;7:1089-1099.

[14] Rasmussen SE, Frederiksen H, Struntze Krogholm K, Poulsen L. Dietary proanthocyanidins: occurrence, dietary intake, bioavailability, and protection against cardiovascular disease. Mol Nutr Food Res 2005;49:159-74.

[15] Hagerman AE. Hydrolyzable tannin structural chemistry. Tannin handbook 2002:1-8.

[16] Matasyoh LG, Murigi HM, Matasyoh JC. Antimicrobial assay and phyto-chemical analysis of *Solanum nigrum* complex growing in Kenya. Afr J Microbiol Res 2014;8:3923-3930.

[17] Akubugwo I, Obasi A, Ginika S. Nutritional potential of the leaves and seeds of black nightshade-*Solanum nigrum* L. Var virginicum from Afikpo-Nigeria. Pak J Nutr 2007;6:323-326.

[18] Huang HC, Syu KY, Lin JK. Chemical composition of *Solanum nigrum* linn extract and induction of autophagy by leaf water extract and its major flavonoids in AU565 breast cancer cells. J Agric Food Chem 2010;58:8699-708.

[19] Price CW, Ronald E. Anthocyanin pigments of Royal Okanogan huckleberry juice. J Food Sci 1995;60:369-374.

[20] Andersen ØM, Opheim S, Aksnes DW, Frøystein NÅ. Structure of petanin, an acylated anthocyanin isolated from *Solanum tuberosum*, using homo-and hetero-nuclear two-dimensional nuclear magnetic resonance techniques. Phytochemical Analysis 1991;2:230-236.

[21] Milner SE, Brunton NP, Jones PW, O'Brien NM, Collins SG, Maguire AR. Bioactivities of glycoalkaloids and their aglycones from *Solanum* species. J Agric Food Chem 2011;59:3454-84.

[22] Harborne J. Methods of plant analysis. Phytochemical methods: Springer; 1984, p. 1-36.

[23] Ding X, Zhu F, Yang Y, Li M. Purification, antitumor activity *in vitro* of steroidal glycoalkaloids from black nightshade (*Solanum nigrum* L.). Food Chem 2013;141:1181-6.

[24] Guo S-B, Tian Y, Jian L-Y. Optimization of ethanol extraction process of *Solanum nigrum* Linn. and structural confirmation of its compounds. Asian J Chem 2014;26:4615.

[25] Yoshida K, Yahara S, Saijo R, Murakami K, Tomimatsu T, Nohara T. Changes caused by included enzymes in the constituents of *Solanum nigrum* berries. Chem Pharm Bull (Tokyo) 1987;35:1645-8.

[26] Gu XY, Shen XF, Wang L, Wu ZW, Li F, Chen B, Zhang GL, Wang MK. Bioactive steroidal alkaloids from the fruits of *Solanum nigrum*. Phytochemistry 2018;147:125-131.

[27] Ripperger H, Porzel A. *N*-Hydroxysolasodine from *Solanum robustum*. Phytochemistry 1992;31:1837-1839.

[28] Adesina S, Gbile Z. Steroidal constituents of *Solanum scabrum* subsp. nigericum. Fitoterapia 1984;55:362-3.

[29] Mibei EK, Ojijo N, Karanja SM, Kinyua JK. Phytochemical and antioxidant analysis of methanolic extracts of four African indigenous leafy vegetables. Ann Food Sci Technol 2012;13:37-42.

[30] Jared JJ, Murungi LK, Wesonga J, Torto B. Steroidal glycoalkaloids: chemical defence of edible African nightshades against the tomato red spider mite, *Tetranychus evansi* (Acari: Tetranychidae). Pest Manag Sci 2016;72:828-36.

[31] Nath LR, Gorantla JN, Thulasidasan AK, Vijayakurup V, Shah S, Anwer S, Joseph SM, Antony J, Veena KS, Sundaram S, Marelli UK, Lankalapalli RS, Anto RJ. Evaluation

of uttroside B, a saponin from *Solanum nigrum* Linn, as a promising chemotherapeutic agent against hepatocellular carcinoma. Sci Rep 2016;6:36318.

[32] Bai Y. Simultaneous quantification of four saponins, three alkaloids and three fatty acids in *Solanum nigrum* Linn. by HPLC-ELSD. J Med Plants Res 2012;6.

[33] Wang Y, Xiang L, Yi X, He X. Potential anti-inflammatory steroidal saponins from the berries of *Solanum nigrum* L. (European black nightshade). J Agric Food Chem 2017;65:4262-4272.

[34] Friedman M. Chemistry and anticarcinogenic mechanisms of glycoalkaloids produced by eggplants, potatoes, and tomatoes. J Agric Food Chem 2015;63:3323-37.

[35] Kamga RT, Kouamé C, Atangana AR, Chagomoka T, Ndango R. Nutritional evaluation of five African indigenous vegetables. Journal of Horticultural Research 2013;21:99-106.

[36] Dodok L, Modhir A, Buchtova V, Halasova G, Polaček I. Importance and utilization of amaranth in food industry. Part 2. Composition of amino acids and fatty acids. Food/Nahrung 1997;41:108-110.

[37] Shanker K, Gupta S, Srivastava P, Srivastava SK, Singh SC, Gupta MM. Simultaneous determination of three steroidal glycoalkaloids in *Solanum xanthocarpum* by high performance thin layer chromatography. J Pharm Biomed Anal 2011;54:497-502.

[38] Silva TMS, Camara CA, Freire KRL, Silva TGd, Agra MdF, Bhattacharyya J. Steroidal glycoalkaloids and molluscicidal activity of *Solanum asperum* Rich. fruits. J Braz Chem Soc 2008;19:1048-1052. [In English]

[39] Johnson AR, Carlson EE. Collision-induced dissociation mass spectrometry: a powerful tool for natural product structure elucidation. Anal Chem 2015;87:10668-78.

CHAPTER II. LEAFY PHYTOCHEMICAL IDENTIFICATION

1 Introduction

The leaves of African *Solanum* nightshades are among the high-priority leafy greens consumed in sub-Saharan Africa. The common existence of toxic glycoalkaloids in many *Solanum* species, however, is also associated with the African nightshades, but has not been subjected to thorough investigation. Apart from glycoalkaloids, literature on other phytochemicals in nightshade leaves remain scarce. The aim of this work is to develop LC-MS methods to identify the major phytochemicals including glycoalkaloids in the leafy samples of African nightshades so as to provide further insight into nutrition and toxicity.

2 Materials and Methods

2.1 Chemical reagents

Standard compounds solasodine was purchased from MP Biomedicals (Santa Ana, CA, USA) and solamargine from MedChem Express (Monmouth Junction, NJ, USA). Methanol and HPLC grade water and acetonitrile modified with 0.1 % formic acid were purchased from Fisher Scientific (Fair Lawn, NJ, USA).

2.2 Botanical authentication of the African nightshades

With unusual species and with lesser studied nonmainstream plants, particularly botanicals, medicinal plants and indigenous plants there can be confused as to the species identity. To ensure proper taxonomical descriptions are attributed to the chemical and

biological studies, botanical authentication is a key step in natural products research to ensure that the intended species being studied and/or reported is indeed that species. Botanical authentication begins with proper botanical, chemical and/or genetic taxonomical evaluations of the selected plant to properly and accurately place the plant its correct species. In this dissertation, each of the African nightshades, which include members of different *Solanum* species were intensely taxonomically studied as to their correct species and classified into their respective species by botanists at the World Vegetable Center using traditional botanical (plant morphological characteristics) and genetic analysis to identify the chromosome number or ploidy level in the species. These analyses and studies were conducted by their expert staff and senior scientists at the WorldVeg Center, Arusha, Tanzania and they are among the leading international botanists, curators and plant breeders of African indigenous vegetables including the African nightshades. The germplasm bank and its accessions as maintained by WorldVeg for *Solanum* is available online at: http://seed.worldveg.org/, after which a user can search for each accession and/or specific *Solanum* spp. The classification of the African nightshade accessions including the *Solanum* species was also reported in the detailed review by Yang et al. in 2013 (with one exception in that the species name, *Solanum* sp. as shown on page 150, Figure 4., was only tentatively classified as *S. americanum*, with the botanist indicating a possibility that it may be *S. nigrum* due to the close morphological characteristics). The identification and classification of the African nightshades were also elaborated by Edmonds [1], Lin et al. [2] and Guzman et al. [3]. In our studies, herbarium voucher specimens with the leaves and flowers of the *Solanum* spp. were collected and

reside with the New Use Agriculture and Natural Plant Products Program, Rutgers, the State University of New Jersey.

2.3 Plant collection

Seeds of 15 entries of *Solanum* spp., i.e., *S. nigrum*, *S. scabrum*, *S. americanum* and *S. villosum* (**Table II-1**) were sown under greenhouse conditions at the Rutgers Research Greenhouses in New Brunswick, NJ. After four weeks of growth, the seedlings were transplanted during the first week of June in 2015 into a cultivated field at the Clifford E. & Melda C. Snyder Research and Extension Farm, New Jersey Agricultural Experiment Station of Rutgers University in Pittstown, New Jersey (40.6°N, 75.0°W, 116 m elevation). The leaves of the nightshades were manually harvested with the first harvest occurring 21~28 days post-field transplanting. The aerial parts, leaves and tender stems, were cut ~15 cm above the soil line to allow the plants to regrow for multiple harvesting. The collected aerial parts were dried at 40 C° for two weeks and then ground into powder. The samples were stored in shaded zip-lock bags under room temperature. Equipment

A propane-heated walk-in Powell Maxi Miser tobacco dryer (Bennettsville, SC) was used for sample drying. Agilent 1100 series LC/MSD instrument (Waldbronn, Germany) was used for phytochemical profiling. The HPLC was equipped with an auto-degasser, quaternary pump, thermostatted column compartment and a diode-array detector (DAD). The HPLC-MS interface used an electrospray ionization source (ESI) and the MS featured an ion trap analyzer. The software used was HP ChemStation, Bruker Daltonics 4.1 and DataAnalysis 4.1.

Table II-1. Plant materials of *Solanum* species and origin of different accessions.

Sample code	Species	Source	Source ID
S.n 1	*Solanum. nigrum*	Simlaw Seeds (Kenya)	-
S.n 2		USDA	PI 306400
S.n 3		USDA	PI 312110
S.n 4		USDA	PI 381289
S.n 5		USDA	PI 381290
S.s 1	*Solanum. scabrum*	AVRDC	SS 52
S.s 2		AVRDC	Ex Hai
S.s 3		AVRDC	SS 49; Olevolosi
S.s 4		AVRDC	SS 04.2
S.s 5		AVRDC	BG 16; Nduruma
S.s 6		AVRDC	BG-29
S.s 7		USDA	Grif 14198
S.s 8		USDA	PI 643126
S.a 1	*Solanum. americanum*	USDA	PI 268152
S.v 1	*Solanum. villosum*	*USDA*	*Grif 16939*

S.n: *Solanum nigrum*; S.s: *Solanum scabrum*; S.a: *Solanum americanum*; S.v: *Solanum villosum*

2.4 Sample preparation

2.4.1 Extract without hydrolysis for phytochemical profiling

For phytochemical profiling without hydrolysis, around 200 mg of the sample was accurately weighed and extracted with 25 mL 70 % methanol with 0.1 % formic acid. Each extract was fully vortexed, sonicated in a water bath for 5 min, and then let stand still overnight under room temperature. The extract was centrifuged at 13,000 rpm/min for 10 min and then the supernatant was ready for injection into HPLC-MS.

2.4.2 Extract with hydrolysis for aglycone profiling

Around 200 mg of sample was accurately weighed and hydrolyzed by 20 mL 0.5 M anhydrous methanolic hydrochloric acid in an eight-dram vial with a screw cap, and incubated in a 70 °C water bath for 120 min. After incubation, the hydrolyte was cooled down and basified with 3 mL saturated sodium hydroxide methanolic solution to terminate the hydrolysis reaction, and was then brought to 25 mL by methanol. The final hydrolyte was centrifuged at 13,000 rpm/min for 10 min and then the supernatant was ready for injection into HPLC-MS.

2.5 Phytochemical profiling without hydrolysis

The aerial part extract (without hydrolysis, ***section 2.4.1***) was chemically profiled by HPLC-UV/vis-MS. Column Agilent Polaris 3 Amide C18, 250 × 4.6 mm (Santa Clara, CA, USA) was used for compound separation. HPLC grade water with 0.1 % formic acid was used for mobile phase A and acetonitrile with 0.1 % formic acid for mobile phase B. The gradient started from 10 % to 20% B in 0 to 10 min, 20% to 28 % in 10 to 30 min, 28% to 30% in 30 to 40 min, and 30% to 50 % in 40 to 60 min, followed by 5 min column flushing with 80 % B and another 5 min column equilibration with the starting mobile phase before the next injection. The flow rate was 1 mL/min and the injection volume was 10 μL. The column was kept at 25 °C. The DAD detector was set at 210 nm, 254 nm, 280 nm and 370 nm for signals, and scanning range 200~550 nm with 2.0 nm step for spectrum. About a third of the HPLC eluent was split into the MS detector. In the ESI, the nebulizer needle voltage was set at 3500 V of positive polarity. High purity nitrogen (99.999 %) was used as both nebulizing gas at 40 psi and drying gas at 350 °C with a flow rate of 10 L/min.

High purity helium was used as the collision gas and the collision energy was set at arbitrary compound stability 80 %. The scanning mode was set at positive and the range at 200-1300 *m/z*.

2.6 Aglycone profiling with hydrolysis and MS/MS

Aglycones freed and pooled from corresponding glycosides after hydrolysis (***section 2.4.2***) were profiled for further structural elucidation and identity confirmation. The HPLC-MS method used was the same as that applied for phytochemical profiling without hydrolysis, except the following specification. The column used was Phenomenex Prodigy, 5µ, ODS3, 100 A, 150 × 3.2 mm. The gradient was 25 % to 60 % from 0 to 20min; 60 % to 80 % from 20 to 25 min; 80 % to 90 % from 25 to 30 min and then isocratically held at 90 % until 45 min. Product ion scan (MS/MS) was performed for additional structural elucidation/identity confirmation of the free aglycones, either in separate runs or in time-sectioned manner. Protonated aglycones were selected as the precursor ion and fragments were scanned from 100 to 500 *m/z*. The collision energy was pre-optimized by syringe infusion method using representative aglycone standards for corresponding categories, i.e., fragmentation amplitude of 2.5 was selected for aglycones of flavonols by using quercetin standard, and amplitude of 7.0 for aglycones of glycoalkaloids by using solasodine standard.

2.7 Nomenclature

The nomenclature of flavonoid fragmentation pathway described in prior research [4] was used in this study. The labels $^{i,j}A^{+}$ and $^{i,j}B^{+}$ correspond to the fragmental ions containing A and B ring, respectively, formed by cleavage of the i and j bonds of C ring.

3 Results and Discussion

3.1 Phytochemical identification

Major peaks were identified based on the retention time, UV-Vis spectrum, and MS spectral data, as well as by comparison with authentic standards. A representative chromatogram was shown in **Figure II-1**. The identities, retention time and MS data of all compounds identified in ***section 2.5*** were summarized in **Table II-2**. In the four *Solanum* species, flavonoids and saponins dominate the secondary metabolites portfolio along with phenolic acid chlorogenic acid, occasionally with trace level of glycoalkaloids. Apparent variance in chemical profile was observed within species, as well as striking similarities between them, as semi-quantitatively presented in **Table II-3**.

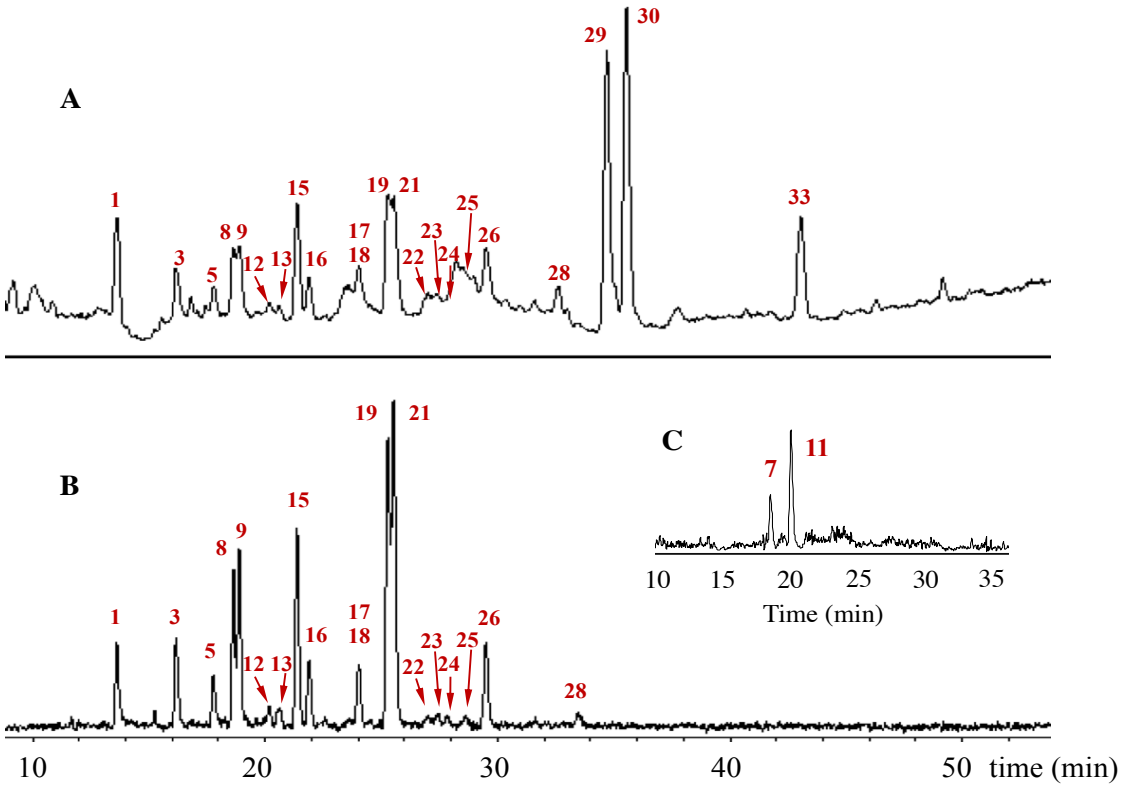


Figure II-1. Representative HPLC-UV/vis-MS chromatograms of *Solanum* spp. (**A**), total ion chromatogram (TIC); (**B**), UV chromatogram at 370 nm, both (A) and (B) acquired from *S. nigrum* PI 312110; (**C**), extracted ion chromatogram (EIC) of glycoalkaloids acquired from *S. villosum* Grif 16939.

Table II-2. Peak assignments used for the analysis of 15 edible nightshade (*Solanum* spp.) accessions

Compound code	RT (min)	Compound ID	$[M+H]^+$ / $[M+Na]^+$ (*m/z*)	Fragment ion (*m/z*)
1	13.6	Chlorogenic acid	*377, 355	-
2	14.0	Quercetin-G-Rha-G	*795	611, 465, 303
3	16.2	Quercetin-G-Xyl-Xyl	*751	597, 465, 303
4	16.8	Quercetin-G-Rha-Rha	*779	611, 465, 303
5	17.8	Quercetin-G-G	*649	465, 303
6	18.1	Quercetin-G-Rha-Xyl	*765	465, 303
7	18.3	Solasodine-G-G-Rha	884	576, 414
8	18.7	Quercetin-G-Xyl	*619	465, 303
9	18.9	Quercetin-G-Xyl	*619	465, 303
10	19.1	Kaempferol -G-G	*633	449, 287
11	20.0	Solasodine-G-Rha-Rha	868	722, 576, 414
12	20.2	Kaempferol-G-Xyl	*603	449, 287
13	20.6	Kaempferol-G-Xyl	*603	287
14	20.7	Kaempferol-G-G	*633	449, 287
15	21.4	Rhamnetin-G-Xyl-Xyl	*765	611, 479, 317
16	21.9	Quercetin-G-Rha	*633	465, 303
17	24.0	Quercetin-G	*487	303
18	24.0	Rhamnetin-G-G	*663	479, 317
19	25.3	Rhamnetin-G-Xyl	*633	479, 317
20	25.5	Kaempferol-G-Xyl	*617	449, 287
21	25.6	Rhamnetin-G-Xyl	*633	479, 317
22	26.9	Rhamnetin-G-Xyl	*633	479, 317
23	27.4	Kaempferide-G-Xyl	*617	301
24	27.8	Kaempferide-G-Xyl	*617	463,301
25	28.6	Rhamnetin-G-Xyl	*633	479, 317
26	29.5	Rhamnetin-G-Rha	*647	479, 317
27	33.3	Tigogenin-G-G-G-G-G	*1249	1087, 903, 741, 579, 417
28	34.4	Tigogenin-G-G-G-Xyl-G	1197	*1057, 903, 741, 579, 417
29	34.8	Dehydrodiosgenin-G-G-Rha-Rha	1029	883, 737, 575, 413
30	35.7	Diosgenin-G-G-Rha-Rha	1031	885, 739, 577, 415
31	41.5	Tigogenin-G-G-Rha-Xyl-Xyl	*1173	887, 741, 579, 417
32	42.6	Tigogenin-G-G-Rha-Xyl-Xyl	*1173	887, 741, 579, 417
33	43.1	Dihydromethyldiosgein -G-Rha-Rha-G	*1069	885, 739, 593, 431
34	45.6	Tigogenin-G-G-G	903	741, 579, 417

*; Sodium adducted ions; RT, retention time; G, glucosyl, galactosyl, mannosyl or other hexosyl; Rha, rhamnosyl; Xyl, Xylosyl.

Table II-3. Phytochemical profile analyses of 15 edible nightshades (*Solanum* spp.) accessions

Peak No.	S. n 1	S. n 2	S. n 3	S. n 4	S. n 5	S. s 1	S. s 2	S. s 3	S. s 4	S. s 5	S. s 6	S. s 7	S. s 8	S. a 1	S. v 1
1	-	T	+	T	T	+	-	+	+	+	+	T	T	+	T
2	-	-	-	-	+	+	-	+	+	T		-	-	++	++
3	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-
4	-	-	-	+	T	++	+	++	+	+	++	+	+		-
5	++	++	+	T	++	+	T	+	+	+	T	T	T	++	++
6	-	-	-	-	-	-	T	+	+	+	T	-	-	-	-
7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T
8	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-
9	-	-	++	-	-	-	-	-	-	-	-	-	-	-	-
10	+	+	-	-	+	-	-	-	-	-	-	-	-	T	-
11	-	-	-	-	T	-	-	-	-	-	-	-	-	-	T
12	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-
13	-	-	T	-	-	-	-	-	-	-	-	-	-	T	-
14	-	-	-	-	-	-	-	T	-	T	-	-	-	-	-
15	-	-	++	-	-	-	-	-	-	-	-	-	-	-	-
16	T	-	+	+	T	+	+	++	+	+	++	+	+	-	T
17	+	T	+	-	+	-	-	-	-	-	-	-	-	T	
18	-	-	+	-	T	-	-	-	-	-	-	-	-	++	+
19	-	-	++	-		-	-	-	-	-	-	-	-	-	-
20	-	-	-	T		T	-	T	T	T	T				-
21	-	-	++	-	-	-	-	-	-	-	-	-	-	-	-
22	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-
23	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-
24	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-
25	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-
26	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-
27	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+
28	++	++	-	++	++	++	++	++	++	++	++	++	++	++	++
29	-	-	++	-	-	-	-	-	-	-		-	-	-	-
30	-	-	++	-	-	-	-	-	-	-		-	-	-	-
31	+	++	-	+	++	++	+	++	++	++	+	+	+	++	+
32	+	+	-	+	++	++	+	++	++	++	+	+	+	+	+
33	-	-	++	-	-	-	-	-	-	-	-	-	-	-	-
34	+	++	-	-	-	-	-	-	-	-	-	-	-	-	+

Plant sample codes refer to **Table II-1**, and the compound codes refer to **Table II-2**. "++", strong peaks defined as intensity of UV-vis over 10 mAU or of extracted ion chromatograph (EIC) by extraction of molecular ions and corresponding fragment ions over 5 × 105; "T", trace level, defined as intensity of UV-vis lower than 1 mAU or EIC intensity lower than 5 × 104; "+", peak intensity between "++" and "T"; "-", peaks not detected.

3.1.1 Identification of polyphenols

Flavonoids were previously reported as an important class of bioactive compounds in nightshades [5]. In this work, a major phenolic acid chlorogenic acid along with an

abundance of flavonols were detected and identified in the four AIVs, including glycosylated derivatives of quercetin, kaempferol, rhamnetin and kaempferide. The individual peaks were identified based on UV-Vis spectrum and MS data.

3.1.1.1 Identification of phenolic acid

For the major phenolic acid, compound **1** (RT 13.6 min) exhibited adduct molecular ions at 377 ($[M+Na]^+$) and 355 *m/z* ($[M+H]^+$) and maximum UV peaks at 245 nm and 328 nm, suggesting a chlorogenic acid, neochlorogenic acid or cryptochlorogenic acid which differ at the esterification site of the quinoyl [6]. This compound was further confirmed to be chlorogenic acid by comparison with authentic standards.

3.1.1.2 Identification of flavonols

For identification of flavonols, for example, compound **8** (RT 18.7 min) featured a sodium adduct molecular ion at 619 *m/z* ($[M+Na]^+$), which was fragmented into 465 *m/z* ($[M\text{-xylosyl}+H]^+$), followed by aglycone ion at 303 *m/z* ($[M\text{-xylosyl-hexosyl}+H]^+$) which might be tentatively identified as peak of protonated quercetin, one of the most prevalent and ubiquitous flavonols in plants. The maximal UV-vis absorption was around 360 nm, in agreement with that of quercetin glycosides in literature [6]. Thus, compound **8** was identified as quercetin conjugated with hexosyl and xylosyl.

In order to further confirm the quercetin aglycone identity, hydrolysis was conducted to release the aglycones and then chromatographically profiled, and peak at ~7.2 min matching a compound mass of 303 Da corresponded to the prior tentatively identified quercetin aglycone (**Figure II-2**). Following this, product ion (PI) scan was conducted across a time window of 6~8 min, with the associated mass spectrum shown in **Figure II-3**.

Fragmental ions labeled in scarlet color, i.e., 229 *m/z*, 247 *m/z*, 257 *m/z*, 275 *m/z* and 285 *m/z*, suggested a flavonol skeleton and C-ring structure. Peaks with green labels, i.e., 137 *m/z* ($^{0,2}B^+$), 153 *m/z* ($^{1,3}A^+$) and 165 *m/z* ($^{0,2}A^+$) arose from the characteristic RDA cleavage of the C-ring, reflecting the substitution pattern of the A- and B-ring [7]. The PI scan supported the quercetin identity. And by comparing the retention time against the reference standard, the aglycone was finally confirmed as quercetin.

The acquired mass spectrum (especially that from the PI scan) of quercetin could be used as important reference for structural elucidation and identification of less known compounds. For example, the aglycone from the two major flavonoid peaks **19** and **21** may be readily identified as methylated quercetin, evidenced by the 14 Da mass shift (317 *m/z* vs. 303 *m/z*) (**Table II-2**) and longer retention time (~12.5 min, labeled by "R") of the free aglycone after hydrolysis (**Figure II-2**) relative to that of quercetin. However, to identify the substitution location of this methyl group, it is necessary to further examine and compare the PI scan mass spectra of this methyl-quercetin with quercetin (**Figure II-4**). Arguably as the most informative characteristics, the co-existence of 137 *m/z* ($^{0,2}B^+$) of both compounds suggested same structure of the B ring, while the 14 Da shift of fragmental ions of methyl-quercetin relative to those of quercetin, i.e., 179 vs. 165 *m/z* ($^{0,2}A^+$), and 167 vs. 153 *m/z* ($^{1,3}A^+$), suggested methylation on the A ring. Considering that the 5-hydroxyl of A ring tends to form intra-molecular hydrogen bonding with the 4-carbonyl of the C ring and substitution on the 5-hydroxyl group is most unlikely, this methyl-quercetin was therefore identified as 7-methyl quercetin, i.e., rhamnetin. And the identity was further confirmed by comparing with the reference standard.

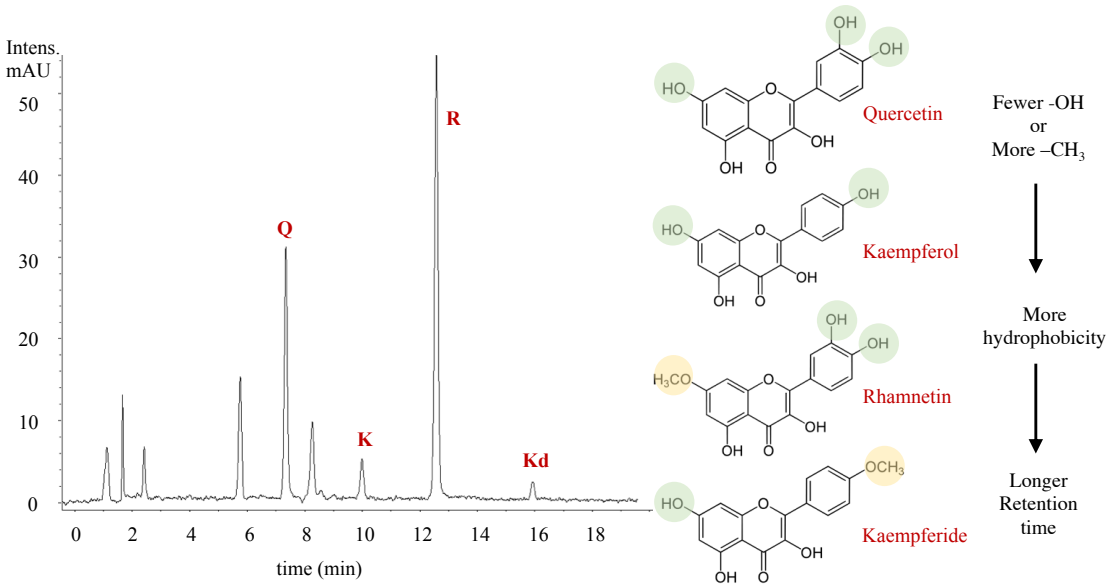


Figure II-2. Representative chromatograph at 370 nm of aerial parts of *Solanum nigrum* PI 312110 after hydrolysis, and the associated aglycone structure comparison. The hydrolysis was performed using anhydrous 0.5 M HCl methanol with 70 °C water bath for 90 min.

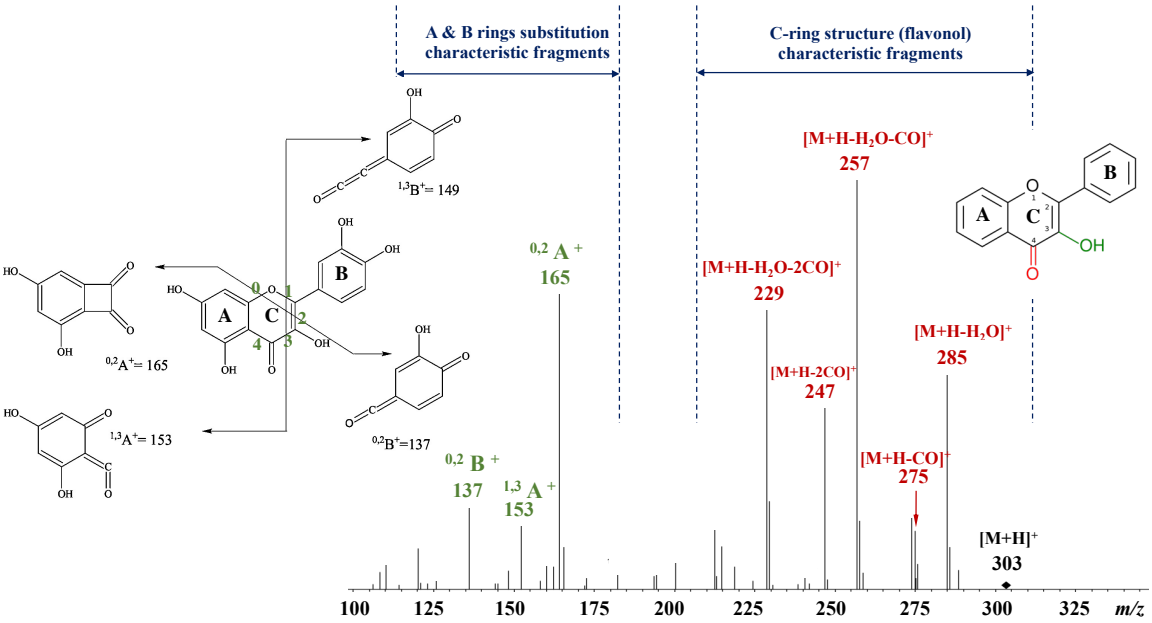


Figure II-3. Representative mass spectra of quercetin as example of flavonol structure elucidation.

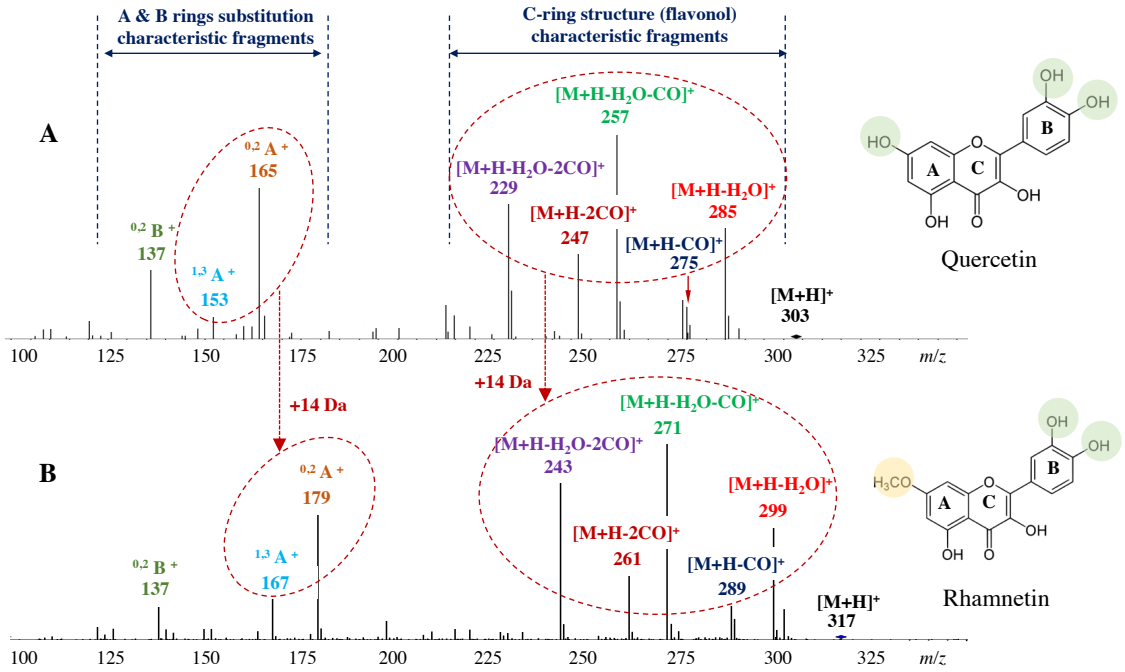


Figure II-4. Mass spectra comparison of quercetin and rhamnetin.

3.1.2 Identification of glycoalkaloids

In this study, no alkaloids were detected in the plant leafy extract except in two accessions, *S.villosum* Grif 16939 and *S.nigrum* PI 381290, where two glycosides of solasodine was found. Compound **11** (RT 20.0 min), for example, has a molecular ion peak at 868 *m/z* $[M+H]^+$, followed by fragment ion 722 *m/z* [M-rhamnosyl+H]$^+$ and 576 *m/z* [M-rhamnosyl-rhamnosyl+H]$^+$, and finally 414 *m/z* [M-rhamnosyl-rhamnosyl-hexosyl+H]$^+$ corresponding to the mass of aglycone solasodine. Thus, compound **11** was identified as solasodine-hexosyl-rhamnosyl-rhamnosyl (solamargine), and the structure was further confirmed as solamargine by comparison with the retention time and mass spectrum of authentic standard. Compound **7** (RT 18.3 min) was identified in like manner and speculated to be solasonine (solasodine-hexosyl- hexosyl-rhamnosyl) as among the most

reported solasodine glycosides in the *S. nigrum* complex. The elution pattern and mass spectra of these two compounds are also in agreement with the literatures [8].

3.1.3 Identification of saponins

Saponin identification was primarily based on mass spectrum interpretation and literature review. Compound **28** (RT 34.4 min), for example, had a molecular ion peak at 1197 *m/z* $[M+H]^+$, which occasionally featured adducted ions at 1219 *m/z* $[M+Na]^+$ and 1237 m/z $[M+Na+H_2O]^+$. The parent ion underwent a loss of hexosyl moiety to generate sodium adducted fragment ion 1057 *m/z* $[M\text{-hexosyl}+Na]^+$, whose remaining saccharide unites were then successively cleaved off to generate the fragment ions at 903 *m/z* $[M\text{-hexosyl-xylosyl}]^+$, 741 *m/z* $[M\text{-hexosyl-xylosyl -hexosyl}+H]^+$, *579* m/z $[M\text{-hexosyl-xylosyl-hexosyl-hexosyl}+H]^+$, and finally the aglycone ion at 417 *m/z* $[M\text{-hexosyl-xylosyl-hexosyl-hexosyl-hexosyl}+H]^+$, which corresponded to the mass of tigogenin as extensively found in *Solanum* complex [9-11]. Thus, compound **28** was identified as tigogenin conjugate with four hexosyls and one xylosyl. The other saponins were identified in similar manner. In addition, the structures of aglycones of diosgenin and tigogenin were further confirmed by comparison with authentic standards after acid-assisted hydrolysis. As glycoalkaloids and saponins are perhaps the most characteristic and interesting phytochemicals in the berries of nightshades, detailed structure elucidation upon these compounds, especially on the aglycone, will be elaborated in the berry chapters.

4 Conclusion

A total of 34 phytochemicals were identified in leaves/aerial parts of African nightshades consisting of four different species, i.e., *S. scabrum*, *S. nigrum*, *S. americanum* and *S. villosum*. The phytochemical insight acquired provided the needed foundation for further quantitative study and associated nutrition and toxicity evaluation.

5 References

[1] Edmonds JM, Chweya JA. Black nightshades: *Solanum nigrum* L. and related species: Bioversity International; 1997.

[2] Lin L, Hsiao Y, Kuo CG. Discovering indigenous treasures: promising indigenous vegetables from around the world: AVRDC-The World Vegetable Center; 2009.

[3] De Guzman C, Siemonsma J. Plant resources of South-East Asia: Backhuys Publ.; 1999.

[4] Ma YL, Li QM, VandenHeuvel H, Claeys M. Characterization of flavone and flavonol aglycones by collision-induced dissociation tandem mass spectrometry. Rapid Commun Mass Spectrom 1997;11:1357-1364. [In English]

[5] Matasyoh LG, Murigi HM, Matasyoh JC. Antimicrobial assay and phyto-chemical analysis of *Solanum nigrum* complex growing in Kenya. Afr J Microbiol Res 2014;8:3923-3930.

[6] Chen HJ, Inbaraj BS, Chen BH. Determination of phenolic acids and flavonoids in Taraxacum formosanum Kitam by liquid chromatography-tandem mass spectrometry coupled with a post-column derivatization technique. Int J Mol Sci 2012;13:260-85.

[7] Tsimogiannis D, Samiotaki M, Panayotou G, Oreopoulou V. Characterization of flavonoid subgroups and hydroxy substitution by HPLC-MS/MS. Molecules 2007;12:593-606.

[8] Jared JJ, Murungi LK, Wesonga J, Torto B. Steroidal glycoalkaloids: chemical defence of edible African nightshades against the tomato red spider mite, *Tetranychus evansi* (Acari: Tetranychidae). Pest Manag Sci 2016;72:828-36.

[9] Vidal Aldana M, Noguiera Lima C. Isolation and characterization of a glycoside from fluid extracts of *Solanum americanum* mill. Afinidad 1999;56:393-396.

[10] Carle R. Investigations on the content of steroidal alkaloids and sapogenins within *Solanum* sect. *Solanum* (= sect. *Morella*)(Solanaceae). Plant Syst Evol 1981;138:61-71.

[11] Bai Y. Simultaneous quantification of four saponins, three alkaloids and three fatty acids in *Solanum nigrum* Linn. by HPLC-ELSD. Journal of Medicinal Plants Research 2012;6.

CHAPTER III. LEAFY PHYTOCHEMICAL QUANTIFICATION

1 Introduction

In Chapter II, bioactive phytochemicals in edible African nightshades (EANs) were identified including glycosides of solasodine, flavonoid glycosides of quercetin, rhamnetin and kaempferol, and saponins of diosgenin and tigogenin, etc. However, the quantitative determination remained unresolved. Quantification of individual phytochemical is difficult due to the large number of naturally occurring glycosides, and high cost and limited availability of required reference standards. A most common alternative is quantification of the corresponding aglycones released free after acid-assisted hydrolysis [1, 2]. Acid-hydrolysis methods for glycosides of flavonoids, glycoalkaloids and saponins are many [1, 3, 4], but a validated simultaneous hydrolysis method for all such different type of compounds identified in EANs, though crucial for rapid quantification, has not been reported. With such simultaneous hydrolysis method, if developed, quantification of freed aglycones could then be readily achieved by application of ultra-high performance liquid chromatography coupled with tandem mass spectrometry (UHPLC-MS/MS) due to its high throughput, selectivity and sensitivity, reliable precision and robustness and large dynamic range of linearity [5].

The aim of this work was to develop a convenient and rapid method for the simultaneous hydrolysis of flavonoid glycosides, glycoalkaloids and saponins in the EANs and to use UHPLC-QqQ-MS/MS methods for quantification of post-hydrolysis aglycones as well as other compounds in EANs of different genetic sources cultivated in varied environment including in Kenya, Africa. From a phytochemical perspective, this work

aimed to further investigate the EANs' nutrition value and anti-nutritive-related consumption safety issue, and to promote the utilization of these EANs as an important additional food supply to improve nutrition and, in many cases, to enhance local economy development by its incorporation as commercial horticultural crops in sub-Sahara Africa.

2 Materials and Methods

2.1 Chemical reagents

Authentic reference standards of chlorogenic acid (1), quercetin (2), kaempferol (3), rhamnetin (6), isoquercetin (9), dioscin (11) were purchased from Sigma-Aldrich (St. Louis, MO), isorhamnetin (4) from Carl Roth (Karlsruhe, Germany), solasodine (5) from MP Biomedicals (Santa Ana, CA), diosgenin (7) from ChromaDex, Inc. (Irvine, CA), tigogenin (8) from ALB Technology Limited (Mongkok Kowloon, Hong Kong, China), and solamargine (10) from MedChem Express (Monmouth Junction, NJ). Methanol, concentrated hydrochloric acid (37%), concentrated sulfuric acid (≥ 98%) and HPLC grade water and acetonitrile with 0.1 % formic acid were purchased from Fisher Scientific (Fair Lawn, NJ).

2.2 Plant materials

Germplasms of a total of thirteen unique EAN accessions were sourced from the U.S. Department of Agriculture (USDA), the World Vegetable Center (WorldVeg) and private seed companies, with additional one from an unbranded package marketed close to the Eldoret, Kenya. These accessions were identified as *S. nigrum* and *S. scabrum,* or not specified. Field trials were performed in Rutgers University (RU) (lat. 40.5°N, long.

74.4°W) and Eldoret, Kenya (lat. 0.63°N, long. 35.0°E), producing a total of twenty differently-sourced EANs (**Table III-1**).

For the RU trial, germplasms were sown in 72-cell trays on May 3, 2016 with growing mix (Fafard Grow Mix 2; Sun Gro Horticulture, Agawam, MA) under greenhouse conditions at the RU Research Greenhouses in New Brunswick, NJ. Seedlings were transplanted on June 7, 2016 to RU Horticulture Research Farm III, New Brunswick, NJ. They were planted in raised beds in single rows with 0.032 mm black plastic mulch, spaced 45 cm within rows in plots 2.1 m long and 1.2 m wide, spaced 1 m between plots and 2 m between plot rows. Plants were hand-watered as needed until established in the field, then irrigated using over-head sprinklers. Soluble 15N-15P-15K fertilizer was applied on June 10, 2016 at a rate of 200 $kg \cdot ha^{-1}$. Leaves and tender stems were harvested on July 8, 2016 and dried using a walk-in tobacco dryer unit (Bennettsville, SC) with propane-heated, forced air set to 40°C, and then ground to powder using a shearing-action mill.

For the Kenya field trial, germplasms were sown in 72-cell trays on November 15, 2016 with growing media (Kekkilä Propagation Media, Amiran Kenya Ltd. Nairobi, Kenya) on benches under polyethylene shade. Seedlings were transplanted on January 19, 2017 to a private field and planted in raised beds without mulch, in double rows, under drip irrigation. Plant spacing was the same as the RU trial. Granular 17N-17P-17K fertilizer was applied to beds prior to transplanting at a rate of 278 $kg \cdot ha^{-1}$. Leaves and tender stems were harvested on March 14, 2017. Samples were dried using a solar dryer at the cultivation site and then ground using a shearing-action mill.

Table III-1. Identification, source, field trial site and content of phytochemicals in African edible nightshades.

Code	*Species*	Source	ID	Trial site	Chlorogenic acid	Quercetin	Kaempferol	Rhamnetin	Isorhamnetin	Solasodine	Diosgenin	Tigogenin
1	*S. scabrum*	WorldVeg	BG 29	Kenya	34.12 ± 0.60	326.43 ± 34.87	35.93 ± 4.75	N.D.	4.85 ± 0.32	N.D.	15.88 ± 1.22	486.76 ± 46.72
2	*S. scabrum*		BG 29	RU	31.62 ± 14.83	161.44 ± 3.84	25.13 ± 5.20	N.D.	2.72 ± 0.24	N.D.	15.29 ± 0.57	438.35 ± 5.16
3	*S. scabrum*		Ex Hai	Kenya	20.69 ± 0.33	306.85 ± 55.90	41.41 ± 7.87	N.D.	3.91 ± 0.41	N.D.	17.71 ± 1.98	498.03 ± 65.75
4	*S. scabrum*		Ex Hai	RU	18.85 ± 3.45	126.85 ± 19.50	20.46 ± 2.71	N.D.	2.79 ± 0.31	N.D.	17.29 ± 1.37	448.93 ± 35.23
5	*S. scabrum*		BG 16	Kenya	72.66 ± 15.05	356.61 ± 36.68	40.57 ± 4.31	N.D.	7.34 ± 0.88	N.D.	15.38 ± 0.63	493.19 ± 14.76
6	*S. scabrum*		BG 16	RU	124.07 ± 13.03	280.56 ± 9.22	41.85 ± 8.94	N.D.	6.69 ± 0.44	N.D.	38.15 ± 0.53	829.87 ± 15.32
7	*S. scabrum*		SS 49	Kenya	72.43 ± 41.38	386.50 ± 24.88	51.64 ± 11.24	N.D.	8.24 ± 0.62	N.D.	14.32 ± 0.71	441.31 ± 23.98
8	*S. scabrum*		SS 49	RU	287.32 ± 53.03	511.06 ± 53.40	66.59 ± 14.82	N.D.	12.14 ± 1.24	N.D.	37.85 ± 2.21	834.35 ± 40.15
9	*S. scabrum*		SS 04.2	Kenya	99.63 ± 15.19	386.51 ± 27.14	51.02 ± 5.03	N.D.	9.21 ± 0.82	N.D.	17.26 ± 0.82	498.75 ± 20.93
10	*S. scabrum*		SS 04.2	RU	213.14 ± 65.19	392.98 ± 25.94	49.13 ± 8.21	N.D.	6.96 ± 0.88	0.11 ± 0.02	31.45 ± 2.52	684.14 ± 48.40
11	*S. scabrum*		SS 52	Kenya	101.45 ± 59.15	364.41 ± 17.04	50.83 ± 1.83	N.D.	7.42 ± 0.78	N.D.	22.12 ± 0.20	596.28 ± 9.98
12	*S. scabrum*		SS 52	RU	77.49 ± 9.24	281.76 ± 19.02	47.65 ± 1.64	N.D.	5.64 ± 0.52	N.D.	22.34 ± 1.43	625.87 ± 26.24
13	*S. nigrum*	USDA	PI 306400	RU	69.84 ± 10.07	205.96 ± 10.58	23.16 ± 8.14	N.D.	2.82 ± 0.31	7.80 ± 4.54	20.76 ± 1.62	785.62 ± 29.95
14	*S. nigrum*		PI 381289	RU	270.14 ± 8.48	404.08 ± 31.50	40.17 ± 14.09	N.D.	13.45 ± 0.86	N.D.	74.35 ± 13.54	521.32 ± 2.45
15	*S. scabrum*		PI 643126	RU	122.02 ± 9.93	151.47 ± 16.79	27.34 ± 0.25	N.D.	3.67 ± 0.34	N.D.	44.74 ± 5.13	607.62 ± 24.80
16	*S. nigrum*		PI 312110	RU	281.89 ± 4.20	131.24 ± 6.40	17.12 ± 8.22	351.29 ± 16.41	1.90 ± 0.04	N.D.	159.80 ± 11.00	6.35 ± 0.69
17	*N.D.*	Kenyan market	-	Kenya	66.87 ± 21.02	379.67 ± 62.41	41.42 ± 13.20	N.D.	6.75 ± 0.78	5.41 ± 3.13	27.09 ± 3.43	793.57 ± 99.27
18	*S. nigrum*	Simlaw Seeds	-	Kenya	77.95 ± 4.05	463.46 ± 37.28	42.97 ± 3.81	N.D.	6.08 ± 0.59	1.54 ± 0.56	20.20 ± 0.73	709.01 ± 40.02
19	*S. nigrum*		-	RU	58.19 ± 12.03	237.36 ± 9.03	19.44 ± 1.05	N.D.	3.71 ± 0.38	21.41 ± 1.82	29.49 ± 5.66	612.47 ± 33.30
20	*S. scabrum*	Baker Creek Heirloom	-	RU	96.30 ± 28.44	243.02 ± 18.09	24.91 ± 5.65	N.D.	5.72 ± 0.48	N.D.	54.76 ± 2.04	388.13 ± 17.68

Unit in mg/100g DW. For abbreviations: N.D., not determined (Species) or not detected (phytochemical content); WorldVeg, World Vegetable Center; USDA, U.S. Department of Agriculture; RU, Rutgers University research farm.

2.3 Instrumentation

Quantitative analysis was conducted using an Agilent 1290 Infinity II ultra-high performance liquid chromatography (UHPLC) hyphenated with 6470 triple quadrupole mass spectrometry (QqQ) (Agilent Technologies, Waldbronn, Germany). The UHPLC was equipped with a built-in auto-degasser, binary pump and column thermostat, with the diode-array detector (DAD) bypassed from flow path. Compound separation was performed using Waters Acquity BEH C18 column, 50 x 2.1 mm, 1.7 μm (Milford, MA) protected with Waters Acquity UPLC BEH C18 VanGuard pre-column 5 x 2.1 mm, 1.7 μm (Milford, MA). The LC-MS interface was electrospray ionization (ESI) with jet stream. Nitrogen was used as nebulizing gas, drying gas, sheath gas and collision gas. The software was MassHunter Workstation LC/MS Data Acquisition B.08.00, Optimizer B.08.00., Qualitative Analysis B.07.00, and Quantitative Analysis B.07.01.

2.4 Hydrolysis condition optimization

Three consecutive single factor experiments were conducted, changing acid concentration, hydrolysis time and temperature one at a time. A mixed stock solution of representative flavonoid glycoside, glycoalkaloid and saponin, i.e., isoquercetin (9) (81.3 μg/mL), solamargine (10) (93.3 μg/mL) and dioscin (11) (113.3 μg/mL), respectively, was prepared for hydrolysis experiment. An aliquot of 0.5 mL stock solution was mixed in a screw-capped 1 dram vial with 0.5 mL methanolic solution of sulfuric acid at varied concentrations (to make a final concentration of 0.2 ~ 5 M), vigorously vortexed, and incubated in a water bath of different temperature (50 ~ 90 °C) for varied time (20 ~ 240 min). After reaction, the solution was chilled in cold water and brought to 25 mL. 70 μL

was then aliquoted and mixed with 1 mL methanol, vortexed and centrifuged prior to LC/MS analysis. Glycosides 9, 10, and 11 and the respective aglycone products, i.e., quercetin (2), solasodine (5) and diosgenin (7) were quantified using UHPLC-MS/MS method (a) (section 2.5). Standard mixture of 2, 5, 7, 9, 10 and 11 with series of 1:2 dilution was prepared for construction of the calibration curve. The hydrolysis recovery was calculated as the percentage of aglycones quantified divided by the theoretical amount.

2.5 UHPLC-MS/MS method (a) for hydrolysis optimization

Method (a) was developed for hydrolysis optimization. Water and acetonitrile both modified with 0.1% formic acid were used as mobile phase A and B, respectively. The flow rate was 0.4 mL/min, and the gradient was 25 % B at 0 min, 50% B at 2.5 min, 90 % B at 2.7 min and held until 5 min. The column was equilibrated with 25% B for 2 min between injections. The injection volume was 3 µL. 3 s needle wash using 70% methanol was performed between injections. The column was thermostatted at 30°C. For ESI settings, drying gas was set at 350°C with a flow rate of 13 L/min, the nebulizer was 30 psi, and the sheath gas was 350 °C at 12 L/min. The capillary voltage was 3500 V and 2500 V for positive and negative polarity, respectively. The nozzle voltage was 1000 V and 2000 V for positive and negative polarity, respectively. Dynamic multiple reaction monitoring (dMRM) was used with 0.5 min scanning window centered around the retention time of each compound. MRM transitions were optimized in injection-based method as presented in a most recent study [6]. Product ions of the most abundance were designated as the quantifier ion, and those of the second and third abundance as the primary and secondary qualifier ion, respectively.

2.6 Plant extract preparation

For each plant sample, three replicates were prepared for quantification of post-hydrolysis aglycones (**2**, **3**, **4**, **5**, **6**, **7** and **8)**. Around 50 mg of plant powder was accurately weighted and simultaneously extracted and hydrolyzed in a screw-capped 8 dram vial suing 25 mL methanolic solution of sulfuric acid under the optimized hydrolysis condition. After hydrolysis, the hydrolysate was chilled in cold water, and 30 μL hydrolysate was mixed with 1.5 mL methanol, vigorously vortexed and centrifuged at 12,000 rpm for 10min. The supernatant was then ready for analysis by UHPLC-MS/MS method (b) (section 2.7). Two additional replicates were separately prepared without hydrolysis for quantification of compound **1**. 50 mg of plant powder were accurately weighted, extracted in 25 mL 50 % methanol with sonication for 15 min. 30 μL extract was then mixed with 1.5 mL 50 % methanol, vigorously vortexed and centrifuged at 12,000 rpm for 10min. The supernatant was then ready for analysis by UHPLC-MS/MS method (c) (section 2.7).

2.7 UHPLC-MS/MS methods (b) and (c) for quantification in AENs

Method (b) was developed for quantification of post-hydrolysis aglycones **2**, **3**, **4**, **5**, **6**, **7** and **8**. The gradient was 30% B at 0 min, 45% B at 2.5 min, sharply increased to 90% B at 2.6 min and then held isocratically until 5.2 min. The column was equilibrated with 30% B for 1.5 min between injections. Eluent from 0 to 0.8 min and from 2.6 to 4.2 min was split to waste. For quantification of compound **1** without hydrolysis, method (c) was used. The gradient was 10% at 0 min, 35% at 1.5 min, and 100% at 1.6 min and held isocratically until 2.5 min. The column was equilibrated with 10% B for 1.5 min between

injections. Eluent from 0 to 0.6 min and after 1.5 min was split to waste. MRM transitions for both methods were optimized in like manner as in method (a). The dwell time for each transition was 20 ms. The other LC-MS parameters remained the same as in method (a).

Methods (b) and (c) were validated in terms of sensitivity, linearity, accuracy and precision. Sensitivity included low limit of detection (LLOD) defined as quantifier ion signal to noise ratio (S/N) at 3:1, low limit of quantification (LLOQ) with quantifier ion S/N at 10:1, as well as low limit of primary qualifier ion quantification (LPQ) with primary qualifier S/N at10:1, and low limit of secondary qualifier ion quantification (LSQ) with the secondary qualifier S/N at 10:1. LPQ and LSQ defined the concentration level where full compound identification was achieved based on the designated quantifier/qualifier ratio. Accuracy was validated by spiking known quantities of standards in pure solvent at LLOQ, middle point of linearity (MP) and high concentration level of quantification (HLQ) and calculated as (detected concentration – theoretical concentration) / theoretical concentration x 100 %. Precision was the standard deviation of repeated injections in the same sequence for intra-batch precision (n=6) and separated sequences (n=3) for inter-batch precision, with intra- and inter-batch precision validated at the same three levels as in accuracy validation.

2.8 Statistics

Comparison of the mean level of phytochemical content between species, i.e., *S. scabrum* vs. *S. nigrum*, and comparison between cultivation sites, i.e., RU research farm vs. Kenya Eldoret farm, were conducted using Scheffé's multiple contrast analysis. Partition of variance for species comparison applied nested analysis of variance (ANOVA)

and for cultivation sites two-way factorial ANOVA [7]. Data analysis and visualization was achieved using R Studio with application of *Tidyverse, ComplexHeatmap* as well as other packages [8, 9]

3 Results and discussion

3.1 Hydrolysis method development

3.1.1 Limit of hydrolysis methods in literature

Methods for hydrolysis of polyphenols, glycoalkaloids and saponins are many [1, 3, 4], but concurrent hydrolysis of all three groups of compounds are limited in literature to date. To achieve simultaneous hydrolysis of all three groups could be challenging as one hydrolysis condition suitable for some compounds may be less effective to others. For example, hydrolysis of flavonoid glycosides is commonly undertaken using acidified methanolic aqueous solution, but the high polarity of solvent used is not compatible with the hydrophobicity of sapogenins, the hydrolysis products of saponins, and therefore necessitates additional labor to extract sapogenins by nonpolar solvents for downstream analysis. In addition, limit of water was noted as a critical condition to achieve efficient hydrolysis of glycoalkaloids [10, 11] (see **Figure A-1**. Notice the significant decrease in hydrolysis efficiency due to 3% water in solvent introduced from concentrated hydrochloric acid (37%) vs. anhydrous condition of the same acid concentration). In order to accommodate the polarity and hydrolysis nature of all compounds, anhydrous hydrogen chloride-acidified methanol was applied for simultaneous hydrolysis of flavonoid glycosides, glycoalkaloids and saponins of EAN berries in a recent study [12]. However, this method suffered from tediousness of preparation of the anhydrous condition [13],

limited solvent shelf life (see **Figure A-2**), high material waste and environmental contamination during solvent preparation, and was not convenient for routine application.

3.1.2 Hydrolysis method optimization

As alternative to anhydrous methanolic hydrochloride, methanol acidified with concentrated sulfuric acid (≥ 98%) instead was applied in this study by exploiting the high proton molarity and low moisture level in acid source (see **Table A-1**) and user-friendly preparation procedure. Isoquercetin, solamargine and dioscin were selected as respective representatives of glycosides of flavonol, glycoalkaloids and saponins in EANs, and the concentration was designed such that it would cover the maximum possible concentration of glycosides in the plant extract to ensure the capacity of the hydrolysis method to be developed. The hydrolysis results were presented in **Figure III-1**. Briefly, aglycones of solamargine and dioscin featured similar bell-shaped recovery curve, possibly due to their resembling structure. Under mild conditions, cleavage of glycosidic bonds remained as a rate-limiting step, and increase in acid concentration, hydrolysis time and temperature led to increase in aglycone recovery. While under excessively intense hydrolysis, both aglycones underwent severe degradation and rapidly decreased the recovery rate. In contrast, isoquercetin were almost fully hydrolyzed into the aglycone even under the mildest hydrolysis condition, possibly due to the fewer saccharide units in side chain, and experienced slow aglycone degradation upon harsh hydrolysis conditions.

The “parent” glycosides were also quantified after hydrolysis, but they provided limited indication of hydrolysis efficacy. As isoquercetin was readily hydrolyzed into the aglycone, and solamargine and dioscin were readily broken into corresponding di- or

mono-saccharide counterparts [13], the quantity of the "parent" glycosides were rapidly diminished even in the mildest hydrolysis condition, and did not reflect the true recovery of aglycones.

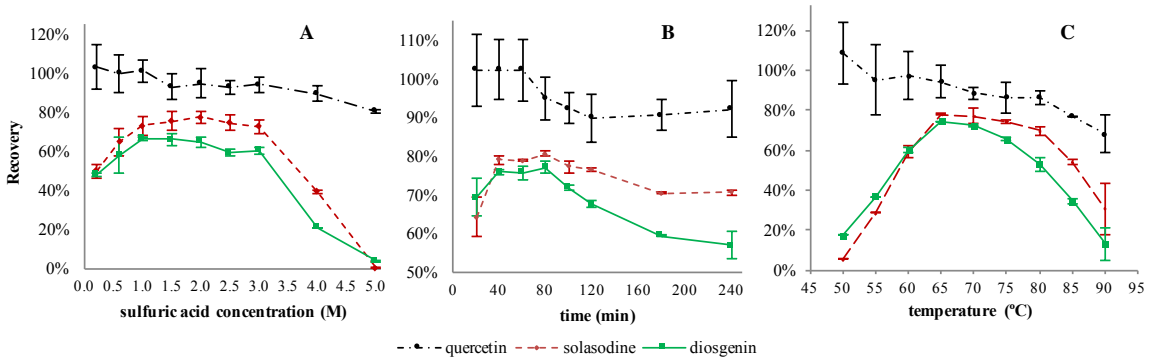


Figure III-1. Recovery of quercetin, solasodine and diosgenin of corresponding glycosides isoquercetin, solamargine and dioscin, respectively, under different hydrolysis conditions. (**A**), 70°C water bath, 90 min; (**B**), 2.0 M sulfuric acid, 70 °C water bath; (**C**), 2.0 M sulfuric acid, 60 min.

The final hydrolysis condition was optimized such that methanol was mixed directly with concentrated sulfuric acid (≥ 98%) to make a final acid concentration of 2M, ideally close to anhydrous condition, with hydrolysis time of 60 min and incubation temperature 65°C. The recovery of quercetin from isoquercetin was 94.4%, solasodine from solamargine 78.1% and diosgenin from dioscin 74.5%. Stability of other aglycones under this hydrolysis condition was also tested, with kaempferol 91.08 %, rhamnetin 73.91 %, and tigogenin 93.90%.

3.2 UHPLC-MS/MS methods

3.2.1 Methods specification

Three different UHPLC-MS/MS methods were developed and tailored to meet different analysis purpose in this study. Method (a) was for quantification of representative glycosides and their corresponding aglycones for hydrolysis optimization. Particularly, the MRMs were updated to dMRM with targeted scanning time windows around the retention time of each compound so as to ensure a sufficient number of data points acquired across chromatographic peaks (see **Figure A-3**. Notice the lack of smoothness in peak curve acquired in MRM instead of dMRM).

Method (b) was for quantification of the post-hydrolysis aglycones. The high hydrophobicity of diosgenin and tigogenin required strong mobile phase composed of up to 90% of acetonitrile for fast elution. Higher proportion of acetonitrile in the mobile phase, however, could cause significant ionization suppression, though higher organic constituent in the mobile phase is generally considered favorable for ionization due to easier solvent evaporation and ion desorption. Solasodine, despite its similar structure to sapogenins, had much weaker retainment on the reverse phase column due to easy ionization of the nitrogenous ring. Considering the wide polarity range of compounds analyzed, to ensure sufficient number of data points acquired across chromatographic peaks (smoothness) and also to reduce instrument contamination, four time-sections were set up in this method. The second section (0.8 ~ 2.6 min) was intended for MRMs of flavonols and solasodine and the fourth section (4.2 ~ 5.2 min) for MRMs of the hydrophobic sapogenins. Eluent of the first section (0 ~ 0.8 min) with unretained "junk" peaks was split to waste. Eluent in the third section (2.6 ~ 4.2 min) during which time 90% acetonitrile was isocratically applied, a typical "column-washing" step, were also split to waste. In addition, in view of the much

higher content of total tigogenin in EANs and its higher sensitivity than most other compounds, the fragmentor voltage for tigogenin was purposely reduced to 0 V instead of 100 V (an optimized voltage yielding the maximum sensitivity) to reduce response to tigogenin and thus to increase the dynamic range of detection of method (b).

Method (c) was set up for quantification of chlorogenic acid without hydrolysis. Due to high polarity of chlorogenic acid, 50% methanol was used as sample solvent instead of 100% methanol in order to avoid compound self-elution in the injection band and thus to ensure good peak shape [14].

3.2.2 MRM fragmentation pathway

Most MRM transitions optimized in this study featured well defined fragmentation pathway. For chlorogenic acid, the 191 *m/z* fragmental ion came from the deprotonated quinic acid moiety (**Figure III-2-D**). Interestingly, this was the single product ion detected, though more than one product ion was readily detectable for isomers neochlorogenic and cryptochlorogenic acid, as also noted in prior research [15]. For flavonol compounds, the intense product ions of 151 *m/z* from quercetin and isorhamnetin and 165 *m/z* from rhamnetin, commonly noted as $^{1,3}A^-$ fragmental ion (containing intact A ring, formed by rupture of bond No.1 and No. 3 in the C ring), all came from the characteristic RDA fragmentation (**Figure III-2-E**) [16-18]. Shared products of 300 *m/z* of rhamnetin and isorhamnetin came from the loss of the methyl group. While positive polarity has been commonly used for flavonols [16, 19], it produced remarkably weaker response than negative polarity in this study (see **Figure A-4**).

For the spirostane-containing compounds, i.e., solasodine, diosgenin, tigogenin, many product ions were produced by the same fragmentation pathway (**Figure III-2-F**). The product ions of 271 *m/z* and 273 *m/z* of diosgenin and tigogenin, respectively, were A-B-C-D ring-containing oxygenous fragments produced through the rupture of the E ring. Product ions of 253 *m/z* from solasodine and diosgenin and those of 255 *m/z* from tigogenin both were deoxygenated A-B-C-D rings formed by respective dehydration from products of 271 *m/z* and 273 *m/z*. Product ions of 157 *m/z* from both solasodine and diosgenin were C-D ring-containing fragments derived from 253 *m/z* formed by cleavage in the B ring [20, 21]. Such fragmentation pattern was also seen in the corresponding glycosides solamargine and dioscin. In addition, both glycosides were found to have strong in-source fragmentation characterized by direct formation of the aglycone ions in the ESI chamber (see **Figure A-3**). This fragmentation behavior was readily noticeable when method (b) was ran in MRM modes before update to dMRM. As this in-source fragmentation was particularly predominant for dioscin, the aglycone ion was selected as the precursor ion for the quantifier transition.

Table III-2. UHPLC-MS/MS parameters for methods (a), (b) and (c) and validation of methods (b) and (c).

Code	Compound	RT (min)	Polarity	Precursor (*m/z*)	Fragmentor voltage (V)	Quantifier (*m/z*)	CE (eV)	Qualifier 1 (*m/z*)	CE (eV)	ratio (%)	Qualifier 2 (*m/z*)	CE (eV)	Ratio (%)
1	chlorogenic acid	0.98	-	353.1	90	191	13	-	-	-	-	-	-
2	quercetin	1.35 [a] /1.03 [b]	-	301	120	150.9	21	178.9	17	68	121	29	15
3	kaempferol	1.49	-	285	135	93	42	159	34	88	186.9	36	81
4	isorhamnetin	1.56	-	315	120	300.1	22	151	30	13	63.2	54	4
5	solasodine	2.22 [a] /2.04 [b]	+	414.3	160	157	41	253.1	33	91	396.4	33	65
6	rhamnetin	2.18	-	315	120	165	21	300	21	25	121	29	18
7	diosgenin	4.55 [a] /4.47 [b]	+	415.3	130	271.2	17	253.1	25	39	157	41	21
8	tigogenin	4.72	+	417.3	100 ‡	273.2	17	161.1	33	58	255.2	29	23.8
9	isoquercetin	0.58	-	463.1	142	300.1	27	300.9	23	57	271	47	48
10	solamargine	1.05	+	868.5	206	129	52	850.5	56	76	253.5	56	38
11	Dioscin	3.35	+	415.3 [§]	206	271.2	17	-	-	-	-	-	-
			+	869.5	206	-	-	253.1	31	17	146.9	31	16

Code	LLOD (ng/ml)	LLOQ (ng/ml)	LPQ (ng/ml)	LSQ (ng/ml)	linear range (ng/ml)	calibration curve	R^2 *	Accuracy (%)			intra-day ($n = 6$) precision (%)			inter-day ($n = 3$) precision (%)		
								LLOQ	MP	HLQ	LLOQ	MP	HLQ	LLOQ	MP	HLQ
1	0.06	0.12	-	-	0.12 - 996.49	Y=262.11 + 69.82	0.9993	-7.5	3.2	-2.3	25.7	6.2	3.7	18.9	3.4	4.6
2	0.27	0.53	2.12	8.48	0.53 - 542.64	Y=50.53X - 2.79	0.9959	4.6	-3.8	4.7	25.4	11.8	6.8	12.5	3.4	5.9
3	4.95	9.89	19.79	19.79	9.89 - 1266.49	Y=4.517X + 4.30	0.9971	-15.8	12.5	-2.3	10.5	8.9	4.9	9.9	2.6	5.73
4	0.24	0.48	3.85	15.41	0.48 - 986.34	Y=198.61X - 38.26	0.9980	-1.0	2.1	-3.3	19.1	6	5.7	11.9	1.5	5.3
5	0.02	0.049	0.19	0.19	0.05- 404.77	Y=1655.45X + 120.37	0.9969	15	7.9	-4	6.9	3.7	3.3	3.4	3.2	3.1
6	0.32	1.3	2.59	2.59	1.3 - 1326.57	Y=104.52X+ 8.15	0.9967	8.3	1.5	5.1	27.9	10.0	11.0	3.6	3.1	1.42
7	0.10	0.19	0.74	0.74	0.19 - 764.23	Y=511.440- 0.09	0.9973	6.1	-1	-3.4	12.4	1.5	2.3	5.4	1.6	3.3
8	0.20	0.41	0.81	1.62	0.41 - 832.18	Y=243.79 - 26.11	0.9989	-2.8	-3.8	-1.5	9.2	1.4	2.7	3.7	1.5	2

RT, retention time. CE, collision energy. Superscripts "a" and "b" refers to RT by methods (a) and (b), respectively. ‡, Due to the significant amount of tigogenin in plant, the fragmentor voltage was set at 0 V for practical quantification of tigogenin to reduce instrumental sensitivity and avoid potential detector saturation. §, produced by in-source fragmentation as the predominant ion. [$], inter-day precision was conducted over three days. LLOD, low limit of detection; LLOQ, low limit of quantification; LPQ, low limit of primary qualifier ion quantification; LSQ, low limit of secondary qualifier ion quantification; MP, middle point of linearity; HLQ, high concentration level of quantification. *, R^2 was calculated with 1/x weight.

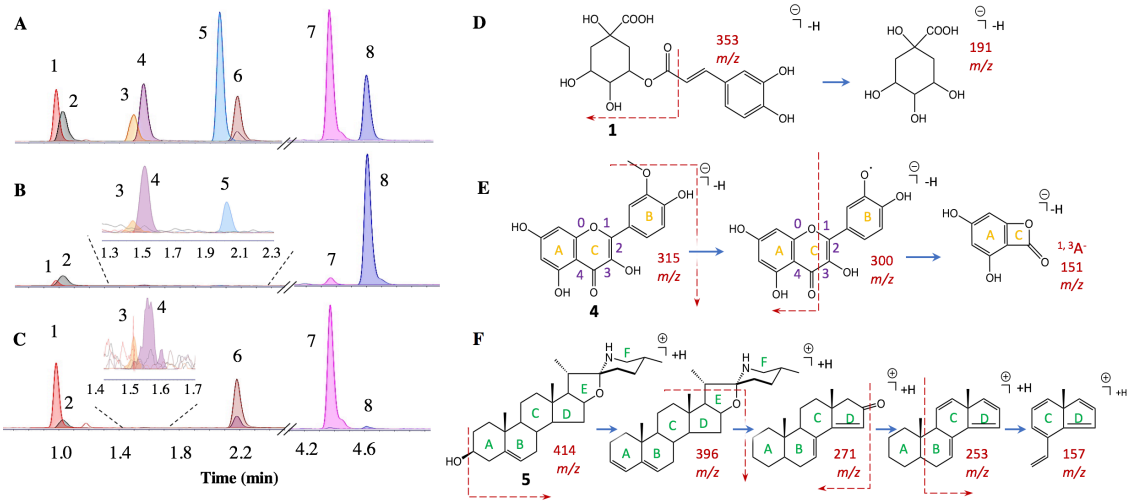


Figure III-2. Representative multiple reaction monitoring (MRM) chromatograms and fragmentation pathways. Chromatographs were shown using quantifier transition for (**A**) standard mixture; (**B**) nightshade sample 17 and (**C**) sample 16. Chromatograms by UHPLC-MS/MS methods (b) and (c) were overlaid. **D**-**F** were to illustrate typical fragmentation pathway the mechanism *per se*, not an exact repetition of MRMs in **Table III-2**. Compound codes refer to **Table III-2**, and sample codes refer to **Table III-1**.

.

3.3 Phytochemical profile of AENs

Methods (b) and (c) were applied for quantification of phytochemicals of EANs from a total of twenty varied sources. Representative chromatograms were exhibited in **Figure III-2**-**B** and **C**. Quantities of phytochemicals were presented in **Table III-1** and visualized as a heat map in **Figure III-3** with sources clustered in hierarchy based on profile similarity. The two species investigated in the RU field trial, i.e., *S. scabrum* (coded samples 2, 4, 6, 8, 10, 12 and 20) and *S. nigrum* (samples 13, 14, 16 and 19), were compared for phytochemical mean content level using Scheffé's multiple contrast analysis (**Figure III-4**-**A1**). Though genetically close and sometimes phenotypically ambiguously distinguished from each other, *S. nigrum* as a group exhibited statistically significant and, intuitively most strikingly and consistently, higher content of solasodine ($p < 0.001$) than *S. scabrum*, which were mostly lacking in such alkaloid. *S. nigrum* also exhibited remarkably more

enriched diosgenin ($p < 0.001$) and tendency for lower content of kaempferol ($p < 0.05$) than *S. scabrum*. While noticing the species difference, it is equally important to notice that within-species variance was tremendous; difference among accessions of the same species contributed to 66 ~ 98 % of total variance across all phytochemicals investigated (**Figure III-4-A2**). The accession from USDA PI 312110 of *S. nigrum* (sample code 16), for example, displayed aberrantly higher content of diosgenin and much lower level of tigogenin (**Figure III-3-A** and **Figure III-4-A1**, also notice the uniquely characteristic rhamnetin) and had great leverage on the statistical significance of the species mean contrast (see **Figure A-5**). Therefore, it is of practical importance to refer to not only the species but also the accession an EAN belongs to for conceptualization of the potential phytochemical content.

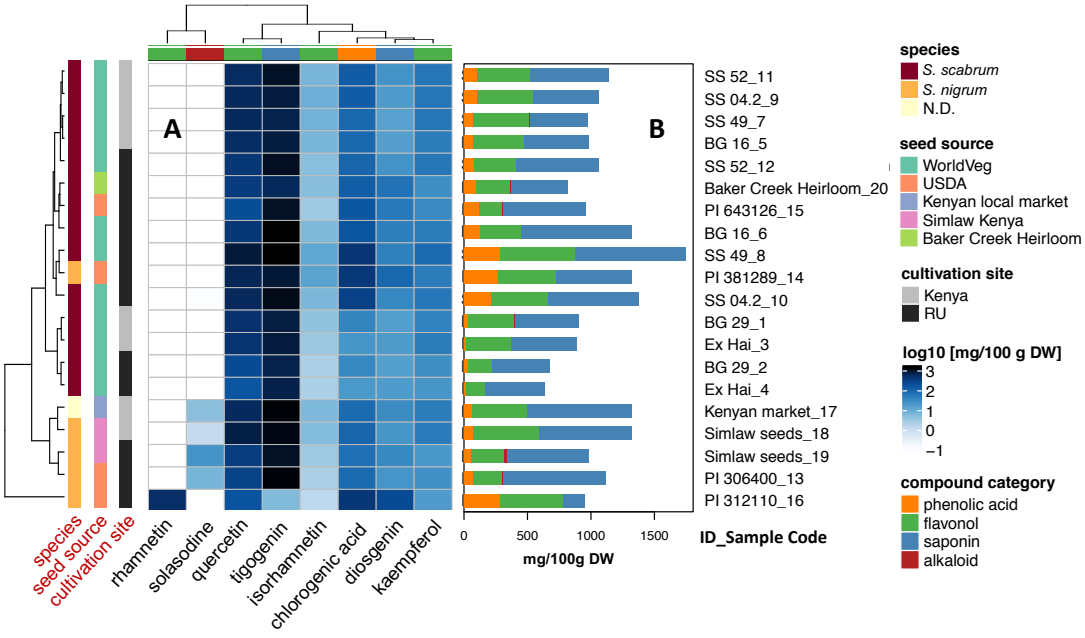


Figure III-3. Phytochemical profile in edible African nightshades from different sources. The clustered heatmap (**A**) shows compound distribution pattern and level of similarity. Stacked barplot (**B**) presents categorized subtotal of compunds in each plan source. Heatmap sidebar and barplot applies the same color notation for compound category.

Apart from difference in species and accessions, the cultivation environment was found to be another important factor influencing the phytochemical profile, as revealed by Scheffé's contrast of the mean of RU field trial (coded samples 2, 4, 6, 8, 10, 12) vs. Kenya Eldoret trial (samples 1, 3, 5, 7, 9, 11), both sites comprising six pairwise accessions of *S. scabrum* (**Figure III-4-B1**). Plants grown in the RU farm showed consistently higher content of both diosgenin and tigogenin ($p < 0.001$) than those in the Kenyan farm, as well as tendency for higher mean in solasodine ($p < 0.001$), though *S. scabrum* as a species was generally deficient in this alkaloid. Across all phytochemicals investigated, 1.6 ~ 34.3 % of total variance could be accounted for by the main effect of environment, and 17.3 ~ 73.6 % when interaction of environment with accessions are considered (**Figure III-4-B2**). Thus, diversity of the environment where the EANs are planted could add to additional complexity to phytochemical content. Despite all the potential sources of variance, all twenty differently-sourced plants presented traits in common. Quercetin was the most abundant polyphenol, followed by chlorogenic acid and kaempferol, accounting for over 98% of total polyphenols, though with the exception of USDA PI 312110 (sample 16) where rhamnetin accounted for 45% of total polyphenols. Isorhamnetin was reported in fruits of S. scabrum in a recent study [22] and for the first time was also found in this work in the leaves of EAN from all twenty sources, though isorhamnetin constituted only a minor percentage of polyphenol portfolio. Both tigogenin and diosgenin were important sapogenins in all twenty sources making up for 0.4~0.8% of dry mass, with tigogenin being the most enriched in content (again except sample 16). The predominance of tigogenin somehow paralleled the pattern reported in an earlier study where out of 32 species in

Solanum section *Solanum*, tigogenin in most species was predominant over diosgenin, if there was any, and only two species had higher content of diosgenin than tigogenin [23].

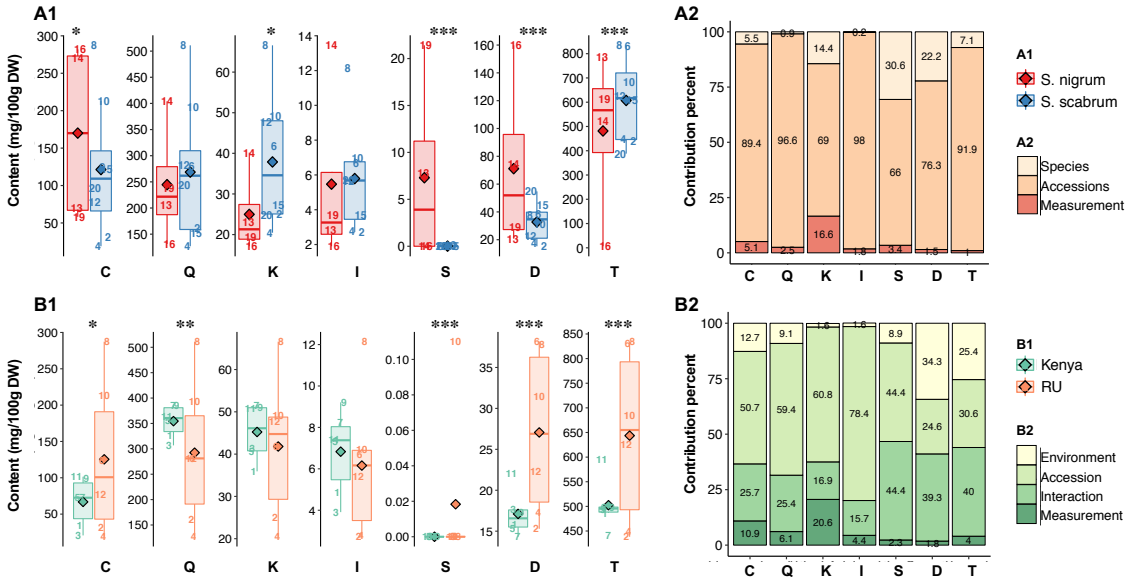


Figure III-4. Contrast analysis and partition of variance between species (**A1** and **A2**) and cultivation environment (**B1** and **B2**). **A1** and **B1** performed Scheffé's contrast for multiple comparison of species mean across different phytochemicals, with statistical significance noted as *, ** and *** at significant levels at 0.05, 0.01, and 0.001, respectively. **A2** performed nested analysis of variance (ANOVA) and **B2** factorial ANOVA. Compounds were abbreviated into initial letters: C, chlorogenic acid; Q, quercetin; K, kaempferol; I, isorhamnetin; S, solasodine; D, diosgenin; T, tigogenin. For boxplot presentation, the arithmetic mean was denoted as the diamond, and numbers indicate the corresponding sample code.

3.4 Nutrition, toxicity and consumption value

Many *Solanum* species are known to contain toxic glycoalkaloids and concern of such toxins in EANs exists to impair their recognition and promotion for food supply. It is of importance to evaluate the potential toxicity associated with the glycoalkaloids, if any in the plants, before consumption. Among the twenty differently sourced EANs, the alkaloid was found in five sources of them with the content ranging between 0.1 to 21 mg/100g dry weight (DW), or an equivalent range of corresponding glycosides of 0.02 ~ 4.2 mg/100 g

fresh weight (FW), a range agreeing with the content reported in prior study in African EANs (0.7 ~ 3.8 mg/100g DW) [11]. The content reported in this study was significantly lower than that of solasodine glycosides (6.25 to 20.5 mg/100 g FW) in eggplant (*Solanum melongena*) that are safe to consume [24, 25]. Therefore, the sufficiently low content of glycoalkaloids in EANs should be within the safety frame for consumption.

Sapogenins as another group of bioactive compounds, though generally to a lesser extent of concern than glycoalkaloids, could also be associated with potential toxicological effects. A recent study concluded safe consumption of 267 mg/kg human body weight (HBW)/day of total steroidal extract comprising saponins of diosgenin and structural analogues [26], or roughly ~133 mg/kg HBW/day of the corresponding aglycone. And another study stated safe consumption of 150 mg/kg animal BW/day of dioscin [27], equivalently a more astringent human dose of 24.2 mg/kg HBW/day [28], , or ~ 11.53 mg/kg HBW/day of the corresponding aglycone. If one of 60 kg were to consume the most sapogenin-dense fresh EANs (~ 872 mg/100g DW or ~ 87.2 mg/100g FW, sample 8) as the only vegetable source and with a daily consumption amount up to 2.5 USA cups (~600g *at maximum* when vegetables were *compressed* tight when measured in cup), a recommended quantity for total dietary vegetables based on 2000 calorie requirement according to the 2015–2020 Dietary Guidelines for Americans [29], the total intake of sapogenins would be 8.72 mg/kg HBW/day, well within the most astringent safety threshold (~11.53 mg sapogenin /kg HBW/day). Therefore, a reasonable consumption amount of EANs would not be likely to cause sapogenin-related health concern.

4 Conclusion

A convenient simultaneous hydrolysis and extraction method for flavonol glycosides, glycoalkaloids and saponins in EANs was developed, applying 2M sulfuric acid in methanol (close to anhydrous condition) with incubation temperature at 65°C for 60 min. A UHPLC-MS/MS method was established to determine post-hydrolysis glycoside-aglycone recovery for hydrolysis method optimization. Two additional UHPLC-MS/MS methods were developed and validated for quantification of eight major aglycones with or without hydrolysis. Difference in species, accessions, and cultivation environment all played important role in affecting the phytochemical profile. Antinutritive alkaloids and sapogenin in EANs were evaluated safe for regular consumption due to lower content than safety threshold, despite concern of toxins commonly associated with *Solanum* species. This work further supported the earlier endeavor in EANs and would facilitate more cultivation, marketing and consumption of EANs to increase food supply, improve household nutrition, and also to provide income generating opportunities for the farmers in the sub-Saharan area.

5 References

[1] Nuutila A, Kammiovirta K, Oksman-Caldentey K-M. Comparison of methods for the hydrolysis of flavonoids and phenolic acids from onion and spinach for HPLC analysis. Food Chem 2002;76:519-525.

[2] Xu Y, Simon JE, Ferruzzi MG, Ho L, Pasinetti GM, Wu Q. Quantification of anthocyanidins in the grapes and grape juice products with acid assisted hydrolysis using LC/MS. J Funct Foods 2012;4:710-717.

[3] Milner SE, Brunton NP, Jones PW, O'Brien NM, Collins SG, Maguire AR. Bioactivities of glycoalkaloids and their aglycones from *Solanum* species. J Agric Food Chem 2011;59:3454-84.

[4] Rothrock J, Hammes P, Mcaleer W. Isolation of diosgenin by acid hydrolysis of saponin. Ind Eng Chem 1957;49:186-188.

[5] Xu RN, Fan L, Rieser MJ, El-Shourbagy TA. Recent advances in high-throughput quantitative bioanalysis by LC-MS/MS. J Pharm Biomed Anal 2007;44:342-55.

[6] Yuan B, Zhao D, Du R, Kshatriya D, Bello NT, Simon JE, Wu Q. A highly sensitive ultra-high performance liquid chromatography/tandem mass spectrometry method with in-source fragmentation for rapid quantification of raspberry ketone. J Food Drug Anal 2018.

[7] Montgomery DC. Design and analysis of experiments: John wiley & sons; 2017.

[8] Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 2016;32:2847-9.

[9] R-Core-Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/. 2013.

[10] Crabbe PG, Fryer C. Modeling of acid hydrolysis of solasodine glycosides. Part III: Combined glycoside hydrolysis and solasodiene formation—guidelines for commercial hydrolysis. AIChE Journal 1983;29:584-587.

[11] Yuan B, Byrnes D, Giurleo D, Villani T, Simon JE, Wu Q. Rapid screening of toxic glycoalkaloids and micronutrients in edible nightshades (*Solanum* spp.). J Food Drug Anal 2018;26:751-760.

[12] Yuan B, Byrnes D, Wu Q, Simon JE. Identification of polyphenols, glycoalkaloids and saponins in *Solanum scabrum* berries using HPLC-UV/Vis-MS J Food Sci 2018.

[13] Crabbe P, Fryer C. Modeling of acid hydrolysis of solasodine glycosides. Part I: Formation of solasodine by glycoside hydrolysis. AIChE journal 1983;29:572-579.

[14] Snyder LR, Kirkland JJ, Dolan JW. Introduction to modern liquid chromatography: John Wiley & Sons; 2011.

[15] Vrhovsek U, Masuero D, Gasperotti M, Franceschi P, Caputi L, Viola R, Mattivi F. A versatile targeted metabolomics method for the rapid quantification of multiple classes of phenolics in fruits and beverages. J Agric Food Chem 2012;60:8831-40.

[16] Tsimogiannis D, Samiotaki M, Panayotou G, Oreopoulou V. Characterization of flavonoid subgroups and hydroxy substitution by HPLC-MS/MS. Molecules 2007;12:593-606.

[17] Ma YL, Li QM, VandenHeuvel H, Claeys M. Characterization of flavone and flavonol aglycones by collision-induced dissociation tandem mass spectrometry. Rapid Commun Mass Spectrom 1997;11:1357-1364. [In English]

[18] Fabre N, Rustan I, de Hoffmann E, Quetin-Leclercq J. Determination of flavone, flavonol, and flavanone aglycones by negative ion liquid chromatography electrospray ion trap mass spectrometry. J Am Soc Mass Spectrom 2001;12:707-15.

[19] Zhen J, Villani TS, Guo Y, Qi Y, Chin K, Pan MH, Ho CT, Simon JE, Wu Q. Phytochemistry, antioxidant capacity, total phenolic content and anti-inflammatory activity of Hibiscus sabdariffa leaves. Food Chem 2016;190:673-680.

[20] Lelario F, Labella C, Napolitano G, Scrano L, Bufo SA. Fragmentation study of major spirosolane-type glycoalkaloids by collision-induced dissociation linear ion trap and infrared multiphoton dissociation Fourier transform ion cyclotron resonance mass spectrometry. Rapid Commun Mass Spectrom 2016;30:2395-2406.

[21] Cahill MG, Caprioli G, Vittori S, James KJ. Elucidation of the mass fragmentation pathways of potato glycoalkaloids and aglycons using Orbitrap mass spectrometry. J Mass Spectrom 2010;45:1019-25.

[22] Yuan B, Byrnes D, Simon JE, Wu Q. LC/UV/MS Method to Access the Quantity of Polyphenols, Alkaloids and Saponins in *Solanum scabrum* Berries of Different Genetic Sources and Maturity. J Sci Food Agric 2018;(in review).

[23] Carle R. Investigations on the content of steroidal alkaloids and sapogenins within *Solanum* sect. *Solanum* (= sect. *Morella*)(Solanaceae). Plant Syst Evol 1981;138:61-71.

[24] Bajaj K, Kaur G, Chadha M. Glycoalkaloid content and other chemical constituents of the fruits of some eggplant (*Solanum melongena*. L.) varieties. J Plant Foods 1979.

[25] Jones PG, Fenwick GR. The glycoalkaloid content of some edible solanaceous fruits and potato products. Journal of the Science of Food and Agriculture 1981;32:419-421.

[26] Zhang X, Jin M, Tadesse N, Xian L, Zhang H, Wang S, Dang J, Zhang Y, Guo Z, Ito Y. Safety investigation on total steroid saponins extracts from *Dioscorea zingiberensis* C.H. Wright: Sub-acute and chronic toxicity studies on dogs. Regul Toxicol Pharmacol 2017;91:58-67.

[27] Xu T, Zhang S, Zheng L, Yin L, Xu L, Peng J. A 90-day subchronic toxicological assessment of dioscin, a natural steroid saponin, in Sprague-Dawley rats. Food Chem Toxicol 2012;50:1279-87.

[28] Nair AB, Jacob S. A simple practice guide for dose conversion between animals and human. J Basic Clin Pharm 2016;7:27-31.

[29] U.S. Department of Health and Human Services USDoA. 2015–2020 Dietary Guidelines for Americans. December 2015.

CHAPTER IV. LEAFY FREE AMINO ACIDS

1 Introduction

African indigenous vegetables (AIVs) are an important food source in sub-Saharan African, and provides the needed food security and diversity and income generating opportunities for local people [1, 2]. Many AIVs have been shown to be nutrient dense [3, 4], and the total protein content typically ranges from ca 20 ~ 50% and even up to 70% of dry mass [5, 6]. Despite literature abundance on total protein of AIVs, the amino acid (AA) composition has rarely been investigated.

Analytical methods for AA have been subjected to long-history and continuous innovation and improvement. Of the numerous techniques developed, the first milestone was the 1950s-invention of ion-exchange chromatography with post-column derivatization using ninhydrin reagent with detection at 570 nm and 440 nm [7]. This technique was later developed into fully automated AA analyzers since 1960s [8], and remains a most frequently applied and classic method to date. This technique, however, suffers most from its elongated hours-long run time per sample and is being gradually replaced by many other methods [9, 10]. One alternative technique is pre-column derivatization (PreCD) with reverse phase (RP) chromatography. The PreCD reagents, such as o-phthalaldehyde (OPA), fluorenylmethyl chloroformate (FMOC-Cl), phenyl isothiocyanate (PITC), dimethylaminonaphthalene-5-sulphonyl chloride (dansyl-Cl) [11-13], and 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate (AQC) [14-16], etc., typically tag the polar AAs with a hydrophobic chromophore, and provides better retainment and separation on the RP column and feasible ultraviolet-visible light and/or fluorescence detection. In

some applications using mass spectrometry (MS) analysis, the labelling reagents including their isotopic counterparts, such as iTRAQTM (isobaric Tags for Relative and Absolute Quantitation) [17, 18] and its later enhanced version aTRAQTM, etc. [19-21], react with amino acids such that the derivatives provide a characteristic fragmentation in MS to generate the product ion corresponding to the labelling reagent. This technique is a cost-efficient solution to provide isotopically-labeled internal standards for each AAs while using a single reagent. All such derivatized methods while providing many merits also present various difficulties, such as the instability and/or inconsistency of derived products, laborious sample preparation, analysis artifacts and/or system contamination, etc. [14].

Different from the former two types of technique based on AAs derivatization, a third technique involves direct analysis of underivatized AAs on the RP column using ion-pairing reagents, such as various perfluorinated carboxylic acids followed mostly with MS detection. While this derivatization-free technique proves to be a more convenient and also powerful tool, the use of ion-paring reagents in the mobile phase, however, could induce noisy background, system peaks and contamination, ion suppression, and long equilibration time especially when the buffer is not balanced [9, 10, 14, 22].

One more technique for underivatized AA analysis that is gaining increasing popularity in recent years is the hydrophilic interaction chromatography (HILIC) with MS detection. HILIC applies polar stationary phase, such as bare silica or polar bonded phase, and high percent of organic mobile phase for separation of polar and charged compounds. The separation mechanism involves analytes' partition between the bulky organic phase and the thin aqueous layer immobilized along the surface of the stationary phase, and many other effects such as ionic interaction and dipole-dipole interaction, etc. [23, 24]. While

this technique has been extensively used for amino acid analysis [25-31], and presented remarkable analytical progress than many other techniques, literature method still face various challenges including undesirable sensitivity, compromised chromatographic performance, limited linear range and long run-time, etc. A brief review of representative literature methods mentioned above is presented in **Table IV-1**.

To overcome the prior setbacks in literatures, this work developed an improved HILIC-MS/MS method for AAs analysis with significantly enhanced sensitivity, improved linear range and higher throughput. Using the developed method, the free AAs in a total of 544 AIVs were determined. Based on acquired free AAs profile, the AIVs categories were successfully predicted using machine learning methods, and an R-Shiny based online interactive application (*https://boyuan.shinyapps.io/AIV_Classifier/*) was constructed for interactive modelling simulation and classification prediction of unknown samples.

2 Materials and Methods

2.1 Chemical reagents

Concentrated hydrochloric acid (HCl) (ca. 36.5~38%), LC/MS grade formic acid (FA), LC grade water and acetonitrile were purchased from Fisher Scientific (Fair Lawn, NJ). LC/MS grade ammonium formate and ammonium acetate, and AA reference standards as listed in **Table IV-2** were purchased from Sigma-Aldrich (St. Louis, MO).

Table IV-1. Representative LC and/or LC-MS methods for amino acid analysis in literature and comparison with our method.

Derivatization	Mobile phase Ø	Chromatography category / column used ▽	Detection	LLOQ (pg on column)¥	Linear range ‡	Time (min)△	Special note	Reference
PreCD with AQC	(A) 9% Waters AccQ·Tag™ eluent A, 91% pure water, pH adjusted to 4.9 with 10% phosphoric acid; (B), 60% ACN in pure water	RP, Waters Acquity UPLC BEH C18 column	UV at 254 nm	48 ~ 442 pg	-	28	Mobile phase may not be compatible with MS due to involatile acid modifier; Ser and Asn, His and Gln coeluted	Austria, 2011 [16] ; a similar one refers to Netherlands, 2007 [15]
PreCD with AQC	(A) 10% AccQ•Tag Ultra eluent A, 90% water (containing 10% ACN, 6% FA, and AMF); (B) AccQ•Tag Ultra solvent B (ACN)	RP, Waters AccQ•Tag Ultra column	Waters Xevo TQ ESI-MS/MS	1.02 fg ~ 1.06 pg (reported as LLOD)	100 ~ 10000	9.5	One of the seven analytical platforms used by the Arabidopsis Metabolomics Consortium	US, 2012 [14]
PreCD with OPA-3MPA and FMOC, automatic in-loop procedure	(A) 40 mM Na2HPO4 in water, pH at 7.8; (B), 45% methanol, 45% acetonitrile and 10% water	RP, Agilent Zorbax Eclipse AAA column	UV at 338 nm; FL excitation / emission at 340 / 450 nm, and 266 / 305 nm	39 ~ 2228 pg	50 ~ 100	26	1) Mobile phase pH was critical. 2) Use 60% MeOH in water for 60 min to clean the column every 25–30 injections.	China, 2017 [12]; similar ones: USA, 2010 [13]
PreCD with aTRAQ™	(A) water, (B) methanol, both containing 0.1% FA and 0.01% HFBA	IP-RP, AB Sciex C18	AB Sciex API 4000 tandem mass spectrometer MS/MS	100 ~ 1000 pg §	-	25	1) Only suitable for MS/MS. 2) A universal solution providing IS for each amino acid. 3) Isotopically improved from iTRAQ™.	USA, 2011 [19]; Similar ones: Slovenia, 2019 [21]
PreCD with TAHS	(A) water with 0.2% acetic acid; (B), acetonitrile with 0.2% acetic acid	RP, GL Sciences Inertsil ODS-3 / Shiseido Capcell Pack MGIII	Sciex API 4000 triple quadrupole MS/MS with TurboIonSpray interface.	5 ~ 34 fg (reported as LLOD)	40 ~ 4000	30	1) High sensitivity. 2) Less desirable chromatographic separation.	Japan, 2009 [20]

Underivatized	(A) 10mM HFBA in 95% water with 5% ACN; (B) 10 mM HFBA in 95% ACN with 5% water.	IP-RP, Waters Acquity BEH C18	Waters Micromass QTOF (MS^1)	200 fg ~ 260 pg	10 ~ 50	16		Sweden & France, 2015 [22]
Underivatized	(A) 0.5 mM TDFHA in water; (B) 0.5 mM TDFHA in ACN	IP-RP, Waters Acquity BEH C18	Waters Micromass Quattro Premier XE MS/MS	2 ~ 15 pg	-	30	10 min equilibration time between injections	Netherlands, 2009 [10]
Underivatized	(A) 0.5 mM TDFHA in water; (B) ACN	IP-RP, Interchrom C18 Uptisphere	Sciex Applied Biosystems, API 2000 triple quadrupole MS/MS	25 ~ 400 pg	5 ~ 1045	31	1) A total of 76 amino acids of biological interest were analyzed. 2) 15 min equilibration time between injections. 3) Flush column for 1 h with ACN (200 µL/min) every 14 samples. 4) Calibration linear range may be limited for certain AAs.	France, 2005 [9]
Underivatized	(A) water, (B) ACN, both containing 0.1% FA	HILIC, Waters Atlantis HILIC silica (150 × 2.1 mm, 3 µm)	Exactive Orbitrap full scan MS1	6 ~ 770 pg $		6 *	Leu & Ile was barely distinguished.	Turkey, 2012 [25]
Underivatized	(A) water with 10 mM AMF and 0.15% FA, pH 3.0; (B), ACN with 2 mM AMF and 0.15% FA	HILIC, Waters Acquity BEH HILIC Amide (100 × 2.1mm, 1.7 µm)	Waters Xevo TQ triple quadrupole MS/MS	1.36 ~ 588 pg	39 ~ 1360	18	1) Chromatographically well separated Leu & Ile. 2) Unbalanced buffer in the mobile phase required long time equilibration time. 3) Notable tailing peaks and narrow linear range for His, Arg & Lys	China, 2013 [26]
Underivatized	(A) 60% water with 5 mM AMF, pH 3.0, and 40% ACN; (B), 95% ACN, 5% water with 5 mM AMF, pH 3.0	HILIC, Waters Acquity BEH HILIC Amide (150 × 2.1mm, 1.7 µm)	Thermo Accela TSQ Quantum™ Access MAX Triple Quadrupole MS/MS	#	-	15	1) Significant tailing peaks of Asp and His (> 3min baseline peak width)	Greece, 2017 [29]

Underivatized	(A) 90% water with 8 mM AMF and 0.12% FA, 10% ACN; (B), 90% ACN, 10% water with 8 mM AMF and 0.12% FA	HILIC, Waters Acquity BEH HILIC Amide (100 × 2.1mm, 1.7 µm)	AB Sciex 4000 QTRAP LC–MS/MS	100 ~ 2000 pg	40 ~ 400	23	1) 10 min equilibration between injections. 2) Short linear range for certain AAs.	Germany, 2016 [27]
Underivatized	(A) water with 10 mM AMF and 0.2% FA; (B), acetonitrile.	HILIC, Waters XBridge Amide (150 mm × 2.1mm, 2.5 µm)	MS/MS	0.74 ~ 3.12 pg	1000 ~ 5000	20	1) Chromatographically well separated Leu & Ile.	China, 2018 [30]
Underivatized	(A) water with 10 mM AMF and 0.15% FA, pH 3.0; (B) 85% ACN, 15% water with 10 mM AMF and 0.15% FA	HILIC, Waters Acquity BEH HILIC Amide (100 × 2.1mm, 1.7 µm)	Waters Xevo-TQ triple quadrupole MS/MS	< 80 pg	2500 ~ 5000	18		Netherlands, 2016 [28]
Underivatized	(A) water with 0.1% FA; (B), ACN with 0.1% FA	HILIC, Waters Acquity BEH HILIC Amide (100 × 2.1mm, 1.7 µm)	Agilent 6470 triple quadrupole MS/MS	0.3 ~ 12 pg, mostly < 1 pg	128 ~ 8200, mostly *ca* 4000	6		Our Method

For reagent abbreviations, AQC, 6-aminoquinolyl-*N*-hydroxysuccinimidyl carbamate; FMOC, 9-fluorenylmethyl chloroformate; OPA-3MPA, *o*-phthalaldehyde 3-mercaptopropionic acid; HFBA, heptafluorobutyric acid; TDFHA, Tridecafluoroheptanoic acid; TAHS, *p*-*N*, *N*, *N*-trimethylammonioanilyl *N*′-hydroxysuccinimidyl carbamate iodide; AMF, ammonium formate; ACN, acetonitrile; FA, formic acid. For instrumentational method abbreviations, PreCD, pre-column derivatization; RP, reverse phase; IP, ion-paring; HILIC, hydrophilic interaction; UV, ultraviolet detection; FL, fluorescence detection; LLOQ, lower limit of quantification, reported in pure solvent unless otherwise specified. Special reagents in mobile phase and the major chromatography technique were bolded in table. #, LLOQ was based on solid-matrix mass and not comparable with other methods, thus not summarized in this table. "-", not reported. §, 1 µL injection volume assumed. *, the equilibration time between injections was not counted. $, LLOQ determined in apple juice. Ø, Mobile phase composition typically refers to volume ratio. For consistency, phase A refers to the water based phase and B the organic phase (may switch the order reported on HILIC method in some literatures). ▽, Since HILIC technique is more of our interest, the column dimension was also summarized in table but not for other chromatographic techniques. ‡, The calibration linear range is the ratio between upper concentration bound vs. lower bound. ¥, when molarity was reported in literature, an average molecular weight of 100 Da was used for conversion to mass unit. △, the total run time includes equilibration time between injections.

2.2 Plant materials

A total of 544 miscellaneous AIVs were analyzed in this work, comprising four categories, including African nightshades, 139 samples, from two identified species *Solanum scabrum* and *S. villosum*, twelve accessions; amaranth, 143 samples, from four species *Amaranth cruentus*, *A. hypochondriacus*, *A. tricolor* and *A. dubius*, with thirteen accessions; spider plants, 172 samples, from a single species *Cleome gynandra*, with nine accessions; and Ethiopian mustard, 90 samples, from a single species *Brassica carinata*, with nine accessions. Using the exact same genetic strains, the AIV seeds were distributed to different sites where each strain was cultivated and harvested, the sites including Rutgers University Snyder Research and Extension Farm (Pittstown, NJ, USA; lat. 40.6 °N, long. 75.0 °W, 116 m elevation); the World Vegetable Center, Arusha, Tanzania; Turbo demonstration farm, Eldoret, Kenya; and Kenya Agricultural & Livestock Research Organization (KALRO), Kibos Road, Kisumu, Kenya. Samples were cultivated and harvested in years 2015 ~ 2018 and different seasons if applicable. The field trial procedure refers to our prior report [32-35]. A detailed information about samples refer to the supplementary Excel file.

2.3 Instruments

The instrument used was Agilent 1290 Infinity II UHPLC coupled with 6470 triple quadruple mass spectrometry (Santa Clara, CA). The UHPLC was composed of a binary pump with built-in online degasser, temperature-controlled autosampler and column thermostat. The HILIC column used was Waters Acquity BEH Amide, 100 × 2.1 mm, 1.7 μm, with BEH Amide VanGuard pre-column 5 × 2.1 mm, 1.7 μm (Milford, MA). The MS

featured electrospray ionization (ESI) with jet stream. The drying gas, sheath gas and nebulizing gas was supplied using a Parker Balston NitroFlow60NA nitrogen generator (Lancaster, NY). The collision gas was high-purity grade nitrogen from Airgas (Radnor, PA). The instrumental software was Agilent MassHunter Workstation LC/MS Data Acquisition B.08.00, Optimizer B.08.00, Qualitative Analysis B.07.00, and Quantitative Analysis B.07.01.

2.4 Sample preparation

For plant extract preparation, ca. 100 mg of dried plant powder were accurately weighed, extracted with 10 mL 100 mM HCl aqueous solution, vigorously vortexed and sonicated for 10 min, and stored in -20 °C. The samples were thawed and conditioned to room temperature upon analysis, diluted by mixing 10 μL extract with 1 mL 10 mM HCl in 90% acetonitrile, and then centrifuged at 10,000 ×g prior to LC/MS analysis.

For AA standard solutions, standards each of ca. 10 mg were mixed together and then dissolved in 30 mL 0.1 M HCl aqueous solution to make the standard stock solution. The stock solution was diluted with 10 mM HCl in 90% acetonitrile (note to be the same solvent used for plant extract dilution) into serial concentrations to construct the calibration curve. Preparation of other solvents routinely used or tested during method development refers to the supplementary material section Solvent preparation.

2.5 UHPLC-QqQ-MS/MS method development

For chromatographic separation, water with 0.1% formic acid was used as mobile phase A and acetonitrile with 0.1% formic acid as mobile phase B. The gradient was 90%

B at 0 min, 88% B at 2 min, 70% B at 3 min, 60% B at 4 min and isocratically held until 5 min. Column was equilibrated with starting mobile phase for 1 min between injections. The flow rate was 0.5 mL/min. The total run time was 6 min. Eluent before 1.5 min and after 4.8 min was split to waste. The column thermostat was set at 30°C. Autosampler was maintained at 4°C. Injection volume was 0.3 μL. A 3 sec needle wash using 90% acetonitrile was applied between injections.

For the MS condition, the drying gas temperature was 300 °C and flow rate was 12 L/min. The sheath gas was at 300°C and 12 L/min. The nebulizer was at 30 psi. The capillary voltage was 3500 V and nozzle voltage at 1000 V. Positive polarity was used. The MRM transitions were optimize using MassHunter Optimizer as prior reported [36], and shown in **Table IV-2**.

2.6 Method validation

Validation generally followed FDA guidelines and literature with adaption [37-39]. The quality control sample (QCS) was prepared by equal-volume blending of AIV extracts from different species and categories. Aliquots of 0.8 mL QCS were spiked with standard stock solutions at volumes of 1000, 500, 200, 100, 50, 20 and 10 μL, noted as levels A, B, C, D, E, F and G, respectively, with each level in quadruplicate. Aliquots of 10 μL of each spiked QCS, and 10 μL of the original QCS also in quadruplicate, were diluted by respective mixing with 1 mL 10 mM HCl in 90% acetonitrile prior to LC/MS analysis. The analysis accuracy was computed as measured concentration in spiked QCS subtracting the endogenous content (result referred to as the “net gain”), then divided by the known spike

amount. The dilution effect due to spiking was carefully counted in calculation, and standard deviation computation followed the error propagation law.

For determination of the matrix effect, aliquots of 0.8 mL 0.1 M HCl aqueous solution were spiked with amino acids standard stock solution and then diluted in like manner as accuracy validation aforementioned. Matrix effect was calculated as the "net gain" divided by the measured concentration in pure solvent, and standard deviation followed error propagation law.

For analysis of precision, the same samples prepared in neat solvent for matrix effect validation was re-used, yet using only one set of the four replicates. Each sample, at levels from a to f, were injected for four times, and from level to level in completely random order.

Lower limit of detection (LLOD) and quantification (LLOQ) were concentrations where signal to noise ratio was 3:1 and 10:1, respectively. For calibration construction, four sets of calibrators were prepared separately from the same standard stock solution and all calibrators were applied for calibration construction. The calibrator accuracy at each concentration level was defined as the back-calculated concentration using constructed calibration divided by expected level. The calibrator error percent at each concentration level was defined as the standard deviation of the peak areas of the calibrator quadruplicates divided by the average peak area. Simple linear regression of error percentages against calibrator sequence roughly reflected the error induced at each dilution step.

2.7 AIV category prediction with machine learning

Principle component analysis (PCA) and linear discriminant analysis (LDA) were applied as exploratory tool to investigate the effect of AIV variety on the profile of AA profile. Based on that, machine learning techniques were applied for AIV category classification prediction, using LDA, quadratic discriminant analysis (QDA), elastic net-regularized logistic regression, random forest (RF), support vector machine (SVM) and naïve Bayes (NB), the latter used as a benchmark of model prediction performance. In addition, the prediction result of each model was ensembled counting the most voted category sample-wisely. The testing set was acquired from 70% of the entire dataset via stratified sampling based on category and cultivation site, and was standardized into z-scores. The remaining 30% was used as the testing set to evaluate model performance, and was standardized based on the mean vectors and covariance matrix of the testing set [40, 41]. All statistical computation and visualization was conducted using R [42, 43].

An online interactive ML application was constructed using R Shiny, referring to *https://boyuan.shinyapps.io/AIV_Classifier/*. This application functions to achieve 1) simulation of the model training and testing procedure, 2) classification prediction of a *single* AIV sample via slider bar-mediated manual feature input, and 3) classification prediction of a *batch* sample via Excel file input. The R script for the online application and all computational analysis and visualization refers to *https://yuanbofaith.github.io/AfricanVegetables_AminoAcids/index.html*.

3 Results and Discussion

3.1 UHPLC-QqQ-MS/MS method development

3.1.1 Optimization of mobile phase

Method development of most LC-MS applications is typically centered around the MS part the workhorse of compound detection. while when HILIC is applied, the special feature of the column often renders development of an optimal chromatographic system a most critical part, which would greatly impact the downstream MS performance. In this work, the chromatographic system was carefully optimized. Two columns Waters BEH HILIC vs. BEH Amide of the same dimension, and two mobile phase buffer salt, ammonium formate (AMF) vs. ammonium acetate both at 10 mM, were tested using 2 × 2 factorial design, and compared in terms of resulted peak shape, signal response, compound retention and separation efficiency. The BEH Amide with use of AMF as the mobile phase buffer was found to generate better results than otherwise combinations, which agreed with the literature report [26]. The use of AMF at 10 mM in the mobile phase, however, resulted in significant suppression of MS signal response. Particularly, for some difficult AAs such as aspartic acid and histidine, which presented remarkable tailing peaks, morphologically reminiscent of a melted butter smeared across a jagged waffle, the signal could be easily blended into background and the sensitivity was the most compromised. Such challenge was also seen and inadequately resolved in recent publications [26, 27]. While much literature resorted to higher injection concentration to overcome the compromised sensitivity, such resort could easily induce unnecessary contamination and compromise instrumental performance in the long term.

Therefore to further improve sensitivity, a range of lower concentrations of AMF in the mobile phase were tested in this work, attempting for a sweet spot with increased sensitivity and reserved peak shape (**Figure IV-1** and **Figure A-6)**. Concisely, reduction

or removal of AMF efficiently increased signal response (3.7 ~ 74 times higher when without AMF compared with using 5 mM AMF) and decreased column retainment, the latter manifested by reduction in both retention time and peak width. The increase in peak area and reduction of peak width jointly led to much more enhanced sensitivity, an effect particularly beneficial for compounds with tailing peaks and compromised sensitivity. In addition, the increased sensitivity allows for detection of product ions of low abundance that are respectively unique to leucine and isoleucine, for the first time allowing the two isomers to be distinguished using MS regardless of chromatographic coelution. In addition to increased sensitivity, the peak shape remained uncompromised or even improved upon reduction or removal of AMF, despite the stereotype on buffer's general importance upon peak shape maintenance. As such in our finalized method, only 0.1% FA was added as mobile phase modifier without use of any mineral buffer. Such practice also greatly simplifies method setup, avoids unnecessary salt-induced contamination, allows for compatibility with most LC-MS methods, and avoids issues caused by salt crystallization in instrumental components in the long term.

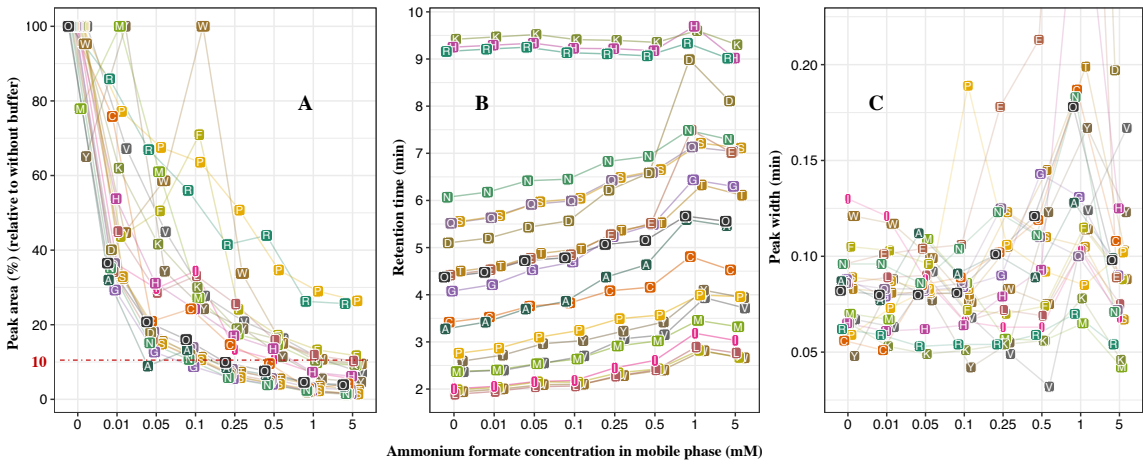


Figure IV-1. Effect of ammonium formate buffer concentration in the mobile phase upon chromatographic performance (CP) of amino acids. The CP metrics include (**A**) signal

response or peak area, (**B**) retention time and (**C**) peak width. A small amount of noise is added to the points' abscissae to reduce overlap. Chromatographic peak shape and more detailed information refer to **Figure A-6**. The formula for mobile phase preparation refers to **Table A-2**. Amino acids' one-letter abbreviations refer to **Table IV-2**.

3.1.2 Optimization of sample preparation solvent

The solvent in the prepared sample ready for LC-MS injection was found as another critical factor for response linearity and sensitivity. Although as a common rule of thumb for HILIC chromatography, the sample solvent remains the same as the starting mobile phase (i.e., 90% acetonitrile with 0.1% or 26.5 mM FA in our case) to ensure undistorted peak shape [44, 45], such solvent surprisingly nonlinearized the signal response of the three basic AAs, histidine, arginine and lysine. When 10 or 100 mM hydrochloric acid was added in replace of FA, the linearity was rapidly restored across three orders of magnitude (**Figure IV-2A** and **Figure A-7**). The sensitivity was also increased compared with using FA. The sample solvent's acid composition did not appear to affect the linear range of other AAs, but also influenced their response sensitivity by various extent. For example, use of HCl at 100 mM in the sample preparation solvent increased sensitivity of tyrosine and methionine, but noticeably reduced sensitivity of glutamic acid, threonine, proline and alanine. Apart from influence on the signal intensity, addition of HCl in the sample preparation solvent also decreased the chromatographic retention time compared with using FA (**Figure IV-2B**). The mechanism of such phenomena is not well understood, but clearly the sample solvent's acid composition changed the existing format of AAs, and more interestingly such format was able to survive through the entire process of column elution during which time the mobile phase poses a very different environment from the sample solvent in the LC vial.

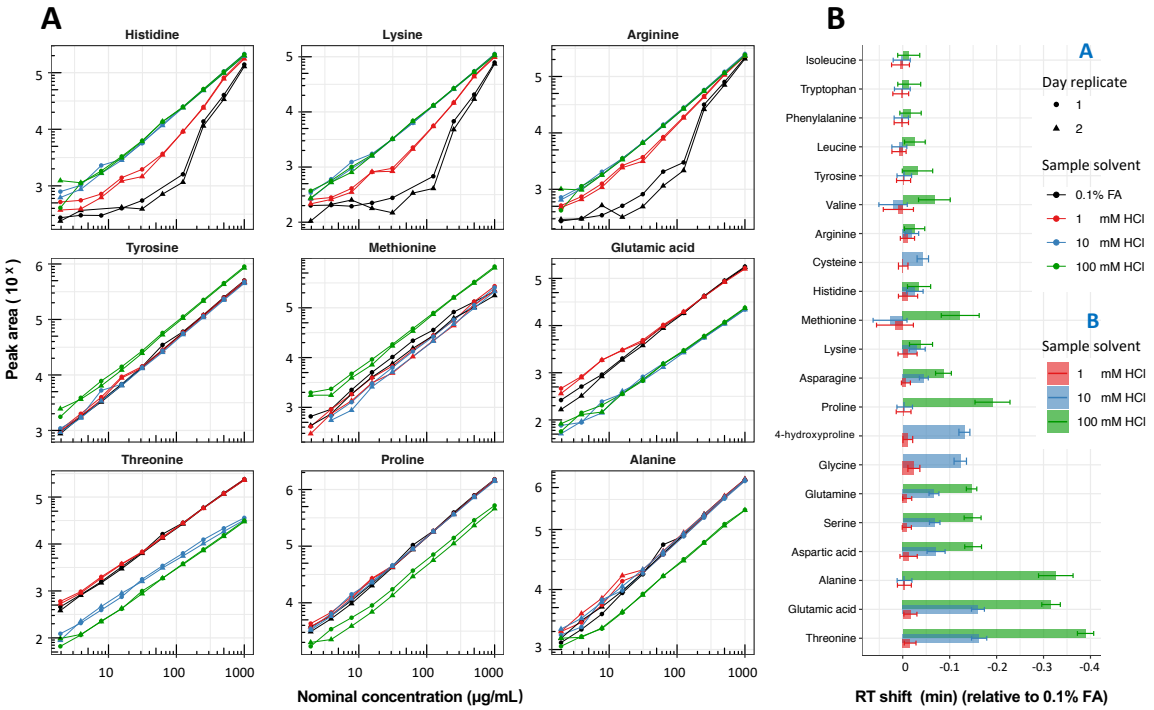


Figure IV-2. Effect of sample preparation solvent on amino acids (AAs) response linearity and retention time. The sample preparation solvents (of the final prepared sample ready for LC-MS injection) were 90% acetonitrile (ACN) with different acid composition, i.e., 0.1% or 26.5 mM formic acid (FA), or hydrochloric acid (HCl) at 1, 10 or 100 mM. In addition, 90 % ACN with 0.1% FA was used as the starting mobile phase. Injection of each calibration set was replicated over two days, during which time the samples were stored in 4 °C autosampler. (**A**), the response curve of representative AAs. Note that the scales are logarithmically transformed, i.e., the calibration curves $y = ax + b$ is re-written as $\log(y) = \log(ax + b) \approx \log(a) + \log(x)$, with y being the signal response, a the slope which reflects the detection sensitivity, x the concentration and b the y-intercept, a and b regarding the original scale. Since the b term is small, the linear range would maintain much of its linearity after log-transformation, with now the new intercept $\log(a)$ positively related with method sensitivity. (**B**), shift of retention time (RT) when HCl is added in the sample solvent compared with 0.1% FA. Since the RT is generally reduced using HCl relative to FA, the shift amount is noted as negative numbers. Certain AAs such as glycine, 4-hydroxyproline and cysteine, etc., shifted out of the dynamic multiple reaction monitoring time window upon 100 mM HCl and thus not detected.

3.1.3 Optimization of the MS/MS

The MS/MS transitions optimized in this work was mostly consistent with the literature [26-28]. Despite the zwitterionic property, positive polarity was found to give the highest sensitivity. Transitions resulting from the neutral loss of formic acid by 46 *m/z* via rearrangement was the most predominant among the AAs as seen in literature [26]. For two isomers leucine and isoleucine, their unique product ions 43.2 and 69.2 *m/z*, respectively, were used as the quantifier ion despite their lower abundance than other product ions. This approach allowed for accurate quantification in presence of chromatographic coelution. For aspartic acid, the lower-abundance product ion 43.2 *m/z* was used as quantifier ion, since the other two product ions 88.1 and 74.1 *m/z* could also be generated from asparagine due to isotopic interference. Lysine and glutamine featured identical transitions, but were sufficiently resolved chromatographically.

In addition, an attempt was made on optimization of the ESI condition using fractional factorial design as prior reported [36], but the signal response was found generally insensitive to different ESI settings, which on the other side suggested robustness of the developed method to ESI condition fluctuations.

Table IV-2. The dynamic multiple reaction monitoring (dMRM) parameters for amino acids.

No.	Amino acids	1-letter Abbr.	3-letter Abbr.	RT (min)	Precursor Ion (*m/z*)	Frag. (V)	Quantifier ion (*m/z*)	CE (eV)	Qualifier ion 1(*m/z*)	CE (eV)	ratio	Qualifier ion 2 (*m/z*)	CE (eV)	ratio
1	leucine	Leu	L	1.614	132.1	74	43.2	26	86.1	10	6.4 ~ 11	44.2	26	1.5 ~ 2.4
2	isoleucine	Ile	I	1.72	132.1	78	69.2	18	86.1	10	9.9 ~ 13.9	44.2	26	2.0 ~ 2.9
3	tryptophan	Try	W	1.792	205.1	76	188	6	146.1	18	.4 ~ 0.61	118.1	30	0.30 ~ 0.39
4	phenylalanine	Phe	F	1.873	166.1	78	120.1	14	103.1	30	0.27 ~ 0.33	77.1	46	0.28 ~ 0.38
5	valine	Val	V	2.18	118.1	54	72.2	10	55.2	22	0.40 ~ 0.48	57.2	34	0.08 ~ 0.12
6	methionine	Met	M	2.346	150.1	78	133	6	56.2	18	0.86 ~ 1.17	61.1	26	0.60 ~ 0.75
7	tyrosine	Tyr	Y	2.771	182.1	78	165.1	6	136.1	14	0.72 ~ 1.24	91.1	30	0.57 ~ 0.92
8	proline	Pro	P	2.989	116.1	84	70.1	18	43.2	38	0.04 ~ 0.06	-	-	-
9	alanine	Ala	A	3.111	90.1	30	44.2	10	-	-	-	-	-	-
10	cysteine	Cys	C	3.452	122	64	59.1	26	76.1	14	0.49 ~ 0.84	-	-	-
11	glycine	Gly	G	3.502	76	20	30.3	14	-	-	-	-	-	-
12	glutamic acid	Glu	E	3.617	148.1	74	84.1	18	130	6	0.48 ~ 0.82	56.2	34	0.25 ~ 0.44
13	threonine	Thr	T	3.655	120.1	68	74.2	10	56.2	18	0.53 ~ 1.04	102.1	6	0.32 ~ 0.48
14	hydroxyproline	Hyp	O	3.694	132.1	86	86.1	14	68.1	22	0.48 ~ 0.74	41.2	34	0.31 ~ 0.60
15	glutamine	Gln	Q	3.887	147.1	74	130.1	6	84.1	18	0.82 ~ 1.03	-	-	-
16	serine	Ser	S	3.898	106.1	74	60.2	10	42.2	26	0.41 ~ 0.52	30.3	22	0.13 ~ 0.18
17	aspartic acid	Asp	D	3.908	134	74	43.2	26	74.1	14	2.60 ~ 3.51	88.1	6	2.34 ~ 3.26
18	asparagine	Asn	N	3.997	133.1	70	74.1	17	87.1	5	0.68 ~ 0.88	-	-	-
19	arginine	Arg	R	4.357	175.1	90	70.1	25	60.2	13	0.25 ~ 0.42	-	-	-
20	histidine	His	H	4.431	156.1	92	110.1	14	83.1	26	0.28 ~ 0.42	-	-	-
21	lysine	Lys	K	4.432	147.1	74	84.1	18	130.1	6	0.48 ~ 0.85	-	-	-

Abbr., abbreviations; RT, retention time; Frag, fragmentor voltage; CE, collision energy. For qualifier/quantifier abundance ratios, the ratio range is established based on the calibrators' response. For Leu, Ile and Asp, the low-abundance product ions were specified as quantifiers to avoid integration interference from other compounds, more discussion referring to ***section 3.1.3***.

3.2 UHPLC-QqQ-MS/MS method validation

The validation results are summarized in **Figure IV-3**, **Table IV-3** and **Table A-3**, **Table A-4** and **Table A-5**. The method's average accuracy for all AAs validated at seven spike levels was centered around 105% and within the range of 75~125%. At lower spike levels, though the accuracy average did not present much deviation from the 100% benchmark, the variance significantly increased, especially when the spike levels went much lower than the content in the original unspiked extract. Since the variance was composed, according to error propagation law, by the error of measurement of both spiked and unspiked extract, the contribution of the latter kept increasing at lower spike levels, and thus boosted the overall accuracy variance (see **Figure A-8**). The averaged level of matrix effect was centered *ca*. the 100% benchmark, mostly bound within 90~110%, suggesting little interference effect imposed from the extract background. The corresponding variance was much larger at lower spike levels (F level not shown) due to the same cause as in case of prior accuracy validation. The precision was generally below 2.5%, though at lower levels the precision error may be up to 15%, reflecting increasing difficulty in integration consistency for smaller peaks, a factor also contributing to the increasing variance of accuracy and matrix effect at lower spike levels. Since injection of samples of different concentrations were made in completely random order, the precision validated here also incorporated the carryover effects if any.

The AAs analyzed featured high sensitivity, with LLOD down to 0.1 ~ 3.2 pg injected on column, and LLOQ down to 0.15 ~ 12.6 pg on column. The linearity range typically spanned over three orders of magnitude, from LLOQ up to high limit of quantification (HLOQ) at 1056 ~ 1716 pg on column (see **Figure A-9)**, which suggested satisfactory

linearity. The calibrators accuracy (CA) was also examined at each concentration level (**Figure A-10A**), with most of the CAs bound within 80~120%. The band of CA fanned out when approaching LLOQ due to increasing difficulty of integration consistency for smaller peaks and the effect of error prorogation of serial dilutions. For the latter, since each set of the four replicates of calibrators were separately prepared from the same stock solution, calibrators incrementally accumulated the dilution error and displayed more CA dispersibility down the serial dilution path. Regression of the calibrator diverging effect, represented as the calibrators error percent, against dilution steps therefore reflect the pattern of error accumulation, and the slope suggested *ca*. 0.5% increase of such error induced at each dilution step (**Figure A-10B**).

Table IV-3. Sensitivity and calibration of amino acids.

No.	Amino acids	LLOD (ng/mL)	LLOQ (ng/mL)	HLOQ (ng/mL)	Calibration	R2
1	leucine	0.50	1.00	4080	Y = 81.4774X + 135.3090	0.9990
2	isoleucine	0.31	0.61	5000	Y = 71.1636X + 111.2613	0.9993
3	tryptophan	0.60	1.19	4880	Y = 168.0951X - 8.2806	0.9962
4	phenylalanine	0.61	1.21	4960	Y = 293.9884X + 260.6369	0.9994
5	valine	0.61	1.22	5000	Y = 410.7137X + 920.6263	0.9989
6	methionine	0.56	1.11	4560	Y = 55.3991X - 26.8628	0.9938
7	tyrosine	0.65	1.31	5360	Y = 65.7911X + 47.0506	0.9995
8	proline	2.05	2.05	4200	Y = 246.6109X + 2569.574	0.9976
9	alanine	0.25	0.50	4120	Y = 140.2096X + 746.8086	0.9967
10	cysteine	1.10	2.21	4519	Y = 13.0453 - 14.8464	0.9640
11	glycine	8.13	16.25	4160	Y = 6.5835X + 17.8001	0.9889
12	glutamic acid	1.40	2.79	5720	Y = 21.7154X + 57.0464	0.9839
13	threonine	1.23	2.46	5040	Y = 29.4302X + 140.7630	0.9954
14	hydroxyproline	4.02	8.05	4120	Y = 33.6225X + 9.4106	0.9996
15	glutamine	2.48	4.96	5080	Y = 25.9211X + 4.1991	0.9988
16	serine	2.25	4.49	4600	Y = 36.0282X + 135.8935	0.9974
17	aspartic acid	10.55	42.19	5400	Y = 4.9136X + 30.8616	0.9988
18	asparagine	1.23	2.46	5040	Y = 14.1428X + 27.9812	0.9993
19	arginine	0.97	1.94	3969	Y = 33.1698X + 293.494	0.9986
20	histidine	2.13	4.25	4352	Y = 39.1237X + 302.7673	0.9994
21	lysine	1.72	3.44	3521	Y = 32.6282X + 198.7964	0.9992

LLOD, lower limit of detection; LLOQ, lower limit of quantification; HLOQ, high limit of quantification. The detection limits correspond to 0.3 µL injection volume. LLOQ and HLOQ constitutes the linear calibration range. The calibration regression was computed using 1/x weight, and based on four sets of calibrator replicates, each set prepared separately by serial dilution from the same stock solution. The four sets were injected over a course of three-day continuous analysis in the same sequence of the 544 samples of African indigenous vegetables.

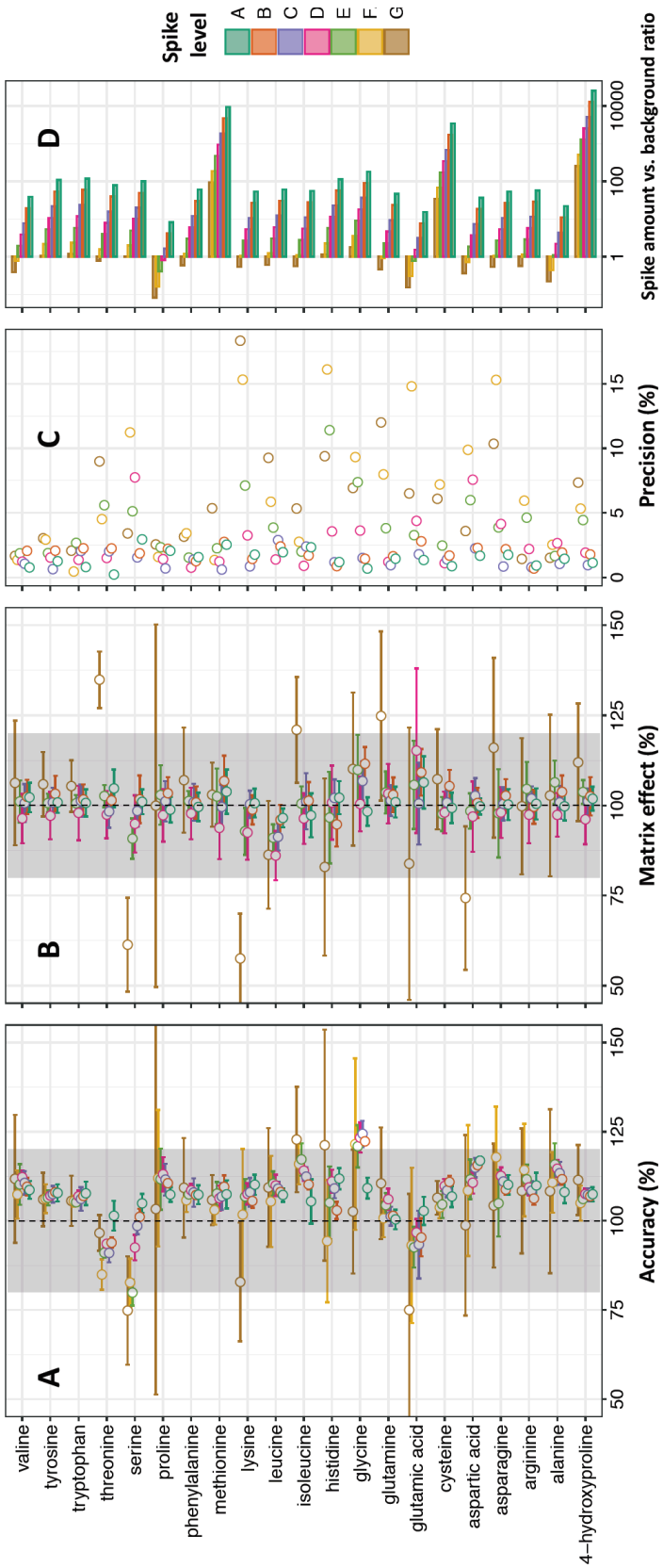


Figure IV-3. Validation of the developed HILIC-QqQ-MS/MS method for amino acids analysis. Spike levels noted as A to G refers to *section 2.6*. Note that the abscissa of plot (**D**) is presented in logarithmic scale.

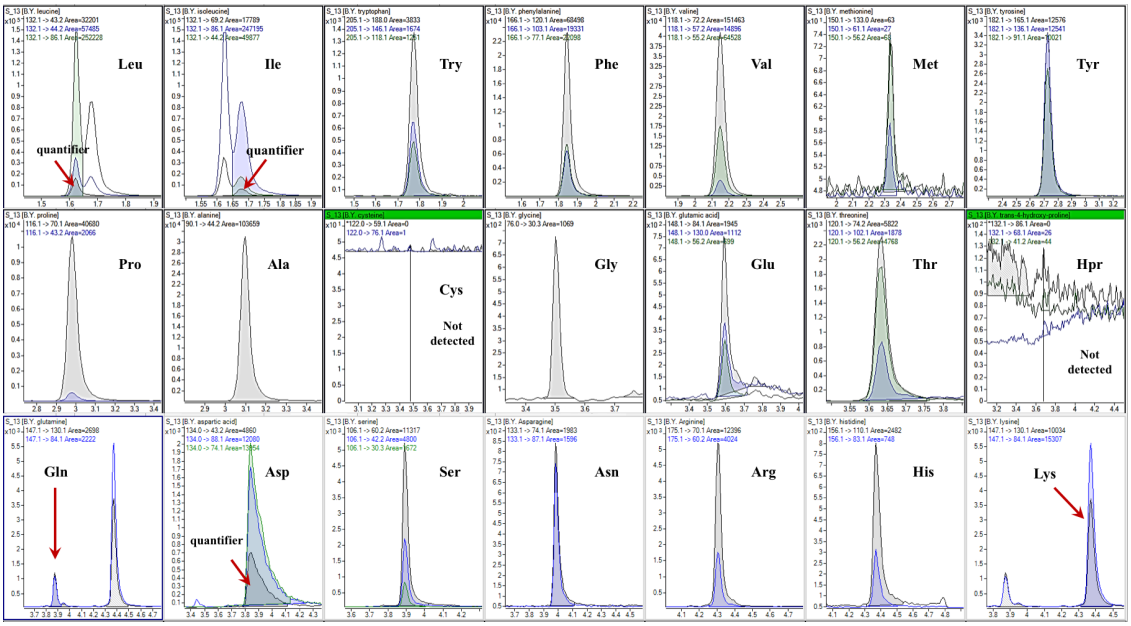


Figure IV-4. Representative chromatogram in a spider plant sample. Compound abbreviations refer to **Table IV-2**. Note that for leucine, isoleucine and aspartic acid, the lower-abundance product ions were used as quantifier ion (see *section 3.1.3*).

3.3 AIVs AA profile

The HILIC UHPLC-QqQ-MS/MS method developed was applied to determine the free AA profile in a total of 544 different AIV samples in a three-day sequence. A representative chromatographic profile of an AIV sample is presented in **Figure IV-4** (prior page). An overview of the profile of all AIVs is presented as a heatmap in **Figure IV-5**. As for total free amino acids, the four categories of AIVs, African nightshades, amaranths, spider plants and mustards, all contained high amount up to 7.4 $\pm$ 3.6 (standard deviation), 2.5 $\pm$ 0.9 , 5.4 $\pm$ 1.7 and 9.3 $\pm$ 2.5 g/100 g dry weight (DW), respectively, which is significantly more enriched than most other vegetables [46]. The high content of total AAs in mustard agreed with the literature, where it was reported in young leaves of

green and red mustards to be 1.13 ~ 1.47 g/100 g fresh weight [47], or equivalently 11.3 ~ 14.7 g/100 g DW assuming 90% moisture content; and the total AAs in Tronchuda cabbage (*Brassica oleracea*) were reported to be 3.3 ~ 14.4 g / kg fresh weight [48], or equivalently *ca*. 3.3 ~ 14.4 g/100g DW.

A particular interest of this work was to investigate the differentiation of AIVs based on the portfolio of free AAs. As such, a PCA was conducted as exploratory analysis for this purpose. Generally, African nightshades, amaranths, spider plants and Ethiopian mustards, were decently separated with each other. Briefly, the amaranth and mustard were nicely separated along the horizontal direction of the first principle component (PC1). Mustards mostly presented positive PC1 scores **(Figure IV-6A)**, and contained higher content of AAs positively correlated with the PC1, such as valines, phenylalanine and histidine, as shown in the loading plot (**Figure IV-6B**). The contributing AAs could also be conveniently manifested by the eigenvector heatmap (**Figure IV-6C**), which presented the coefficients of the linear combination of AAs content for calculation of the PC scores (since AAs content was standardized prior to PCA, for each PC the loadings were also a fixed multiple of the coefficients by the square root of the corresponding eigenvalue). Samples containing more AAs with large PC1 coefficients tend to occupy the right side of the PCA plot, and vice versa. Similar analysis also applied to the separation of nightshades and spider plants in terms of PC2. It's noteworthy that even though the first two PCs only explained 58% of total data variance, the separation was not unclear. Further improved separation was possible when viewed from a higher dimension (see interactive three-dimensional PCA plot at https://yuanbofaith.github.io/aminoAcids_PCA3D/).

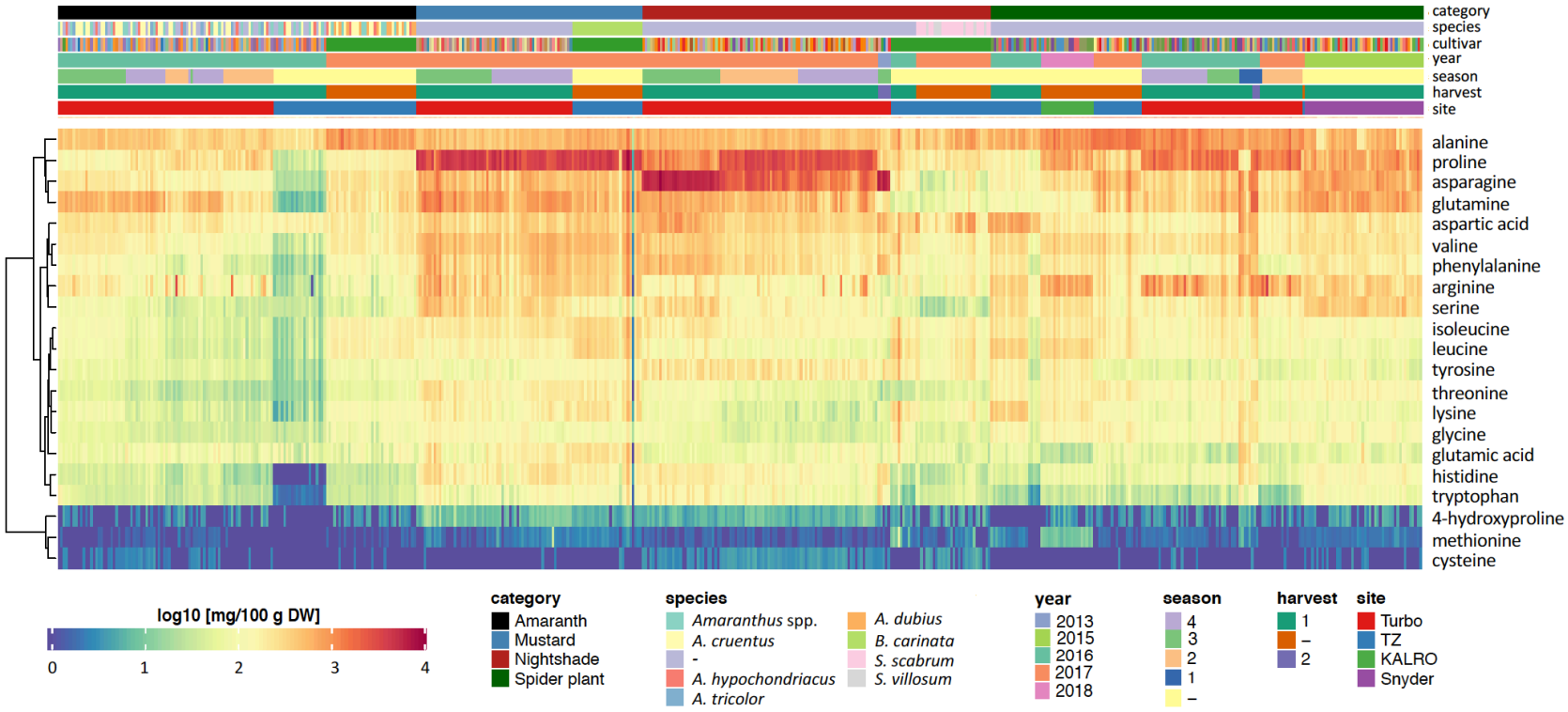


Figure IV-5. Heatmap of free amino acids (AAs) content profile in African indigenous vegetables (AIVs) including a total of 544 samples. The AIV category, species, harvest year, season and site of each individual sample are noted as side bars on top of the heatmap. The names of specific cultivars on the side bar and ID of individual samples on the abscissa is not shown for clarity. Entries noted as "-" are not applicable or recorded. The free AAs' content is presented as mg / 100 g dry weight (DW) with base-ten logarithmic transformation.

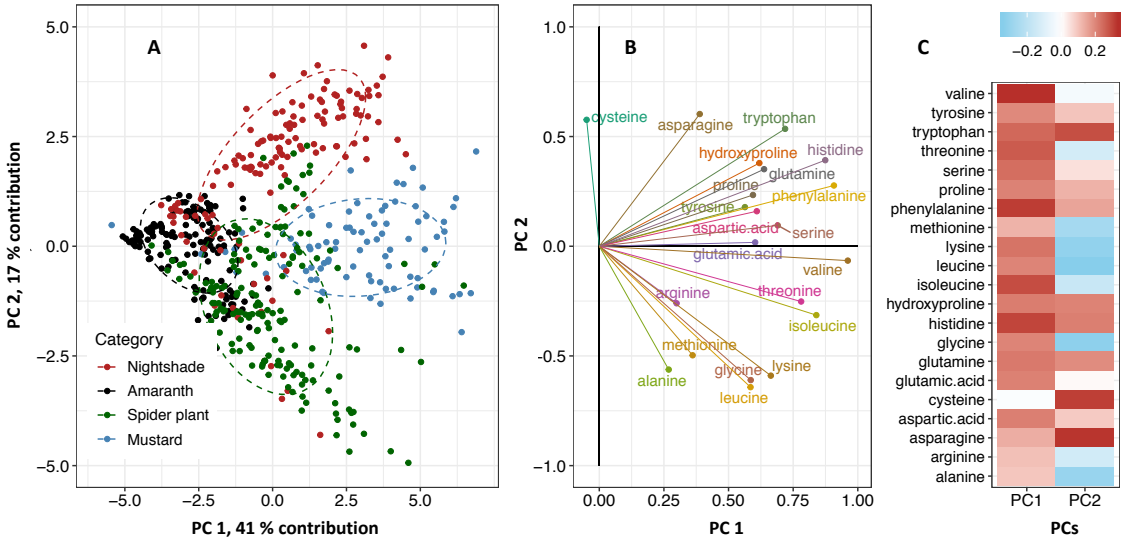


Figure IV-6. Clustering and separation of African indigenous vegetables (AIVs) based on free amino acids (AAs) profile using principle component analysis (PCA). (**A**), the PCA plot of the first two principle components (PC). (**B**), the loadings of the first two PCs, showing the correlation of AAs content with the PCs as well as correlation among AAs *per se*. (**C**), the eigenvector heatmap, showing the linear coefficients used for construction of the first two PCs. An online interactive three-dimensional PCA plot refers to **https://yuanbofaith.github.io/aminoAcids_PCA3D/**.

linear discriminant analysis (LDA) was also applied to linearly maximize the separation (regarding the corresponding mean vectors) among the four AIVs categories. All four categories were decently separated (**Figure IV-7A**). Meanwhile, based on the computed mean vectors, the category of each AIV sample could be conveniently predicted by assigning the sample to the category to which the sample presented the smallest Euclidean distance on the discriminant-valued scale. As such, most samples were correctly predicted with overall 96.1% accuracy (**Figure IV-7B** and **Figure A-11**). This delivers a clear message that using the profile of free AAs, the four categories of AIVs, African nightshades, Ethiopian mustards, amaranth and spider plants, could be predicted with ideal accuracy regardless of the wide within-category variance caused by difference in species, cultivars and cultivation environment.

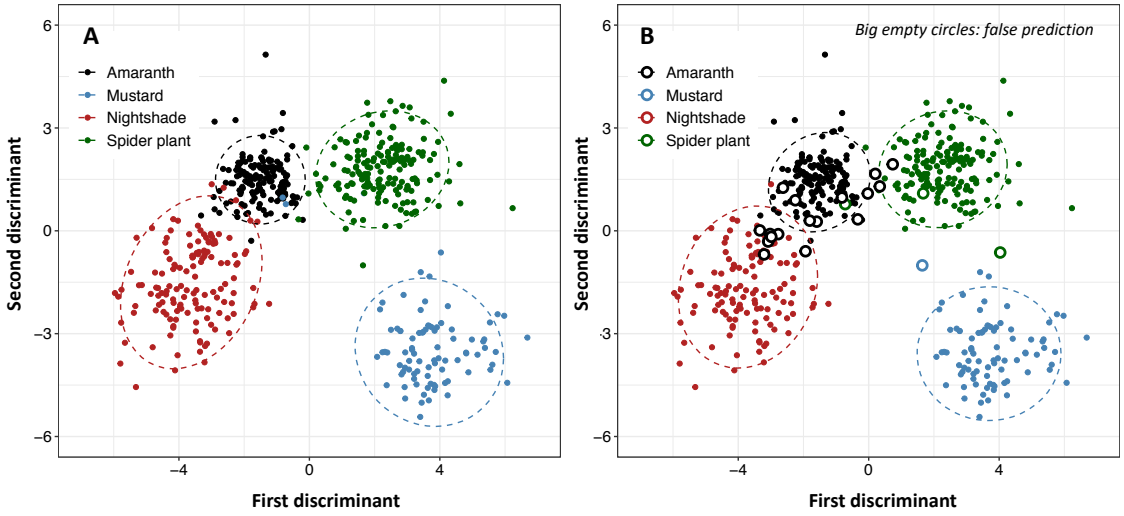


Figure IV-7. Clustering and separation of African indigenous vegetables (AIVs) based on free amino acids (AAs) content using Fisher's linear discriminant analysis (LDA). (**A**), separation of AIVs. (**B**), prediction of AIVs category based on algorithm and statistics of (**A**), with incorrect predictions noted as empty circles. The ellipses mark the boundary containing *ca*. 95% samples corresponding to each category. Note that all 544 AIVs samples are involved in computation, without train-test split.

3.4 AIVs category prediction with ML

Based on prior exploratory analysis, the practice of AIV category prediction, i.e., African nightshades, Ethiopian mustards, spider plants and amaranths, using free AAs' profile was more vigorously tested using different ML techniques. The entire AIVs dataset comprised a total of 544 samples in rows and 21 AAs in columns. The training set contained 70% of the dataset or 381 randomly selected samples. Since the category and cultivation site were two most important factors influencing the AAs profile, stratified sampling was conducted using combinatorial groups of AIV category and cultivation sites. The models were tested on the remaining 30% of the dataset or 163 samples. The result is shown in **Figure IV-8**. The method of NB, with an overall accuracy of 83%, was used as a benchmark of model efficiency considering its "naïve" assumption of the conditional independence among all amino acids, which was clearly untrue (see correlation plot of the

R script documents via the link provided in *section* 2.7). Other ML methods achieved improved accuracy up to *ca*. 95% accuracy. LDA has very quick and simple algorithm, and achieved 93% accuracy. Since LDA assumes equal variance-covariance matrix (VCM) among all four categories and computes the pooled VCM, it does not require a large dataset input to be mathematically feasible. In fact, LDA models trained upon as little as 10% of the original dataset still rendered reasonable prediction result. Such assumption, however, was inaccurate in this work and limited its accuracy. Compared with LDA, QDA computes the VCM separately for each category, more truly reflecting the feature characteristics of each category, and rendered improved accuracy than LDA. Compared with these two discriminant analysis methods which focused on separation of the group means, SVM focuses on boundary optimization and adds more flexibility to nonlinear features. RF with its classification tree-ensembled technique provided quick training and avoided overfitting. EN is a balance between ridge and lasso regression (the alpha parameter set to 0.5 in this work) and provides intuitive model interpretation, though it appeared to be the slowest to train in this work. SVM, RN and EN as well as QDA provided rather similar prediction accuracy, close to 96%.

While the four categories of AIVs were successfully classified using ML methods, prediction on levels of species or even cultivars within each category was not readily achieved. The diverse cultivation environment posed significant influence on AA profile, and such large noise made it difficult to find specific pattern unique to each species and/or cultivar given the limited sample size on such classification levels.

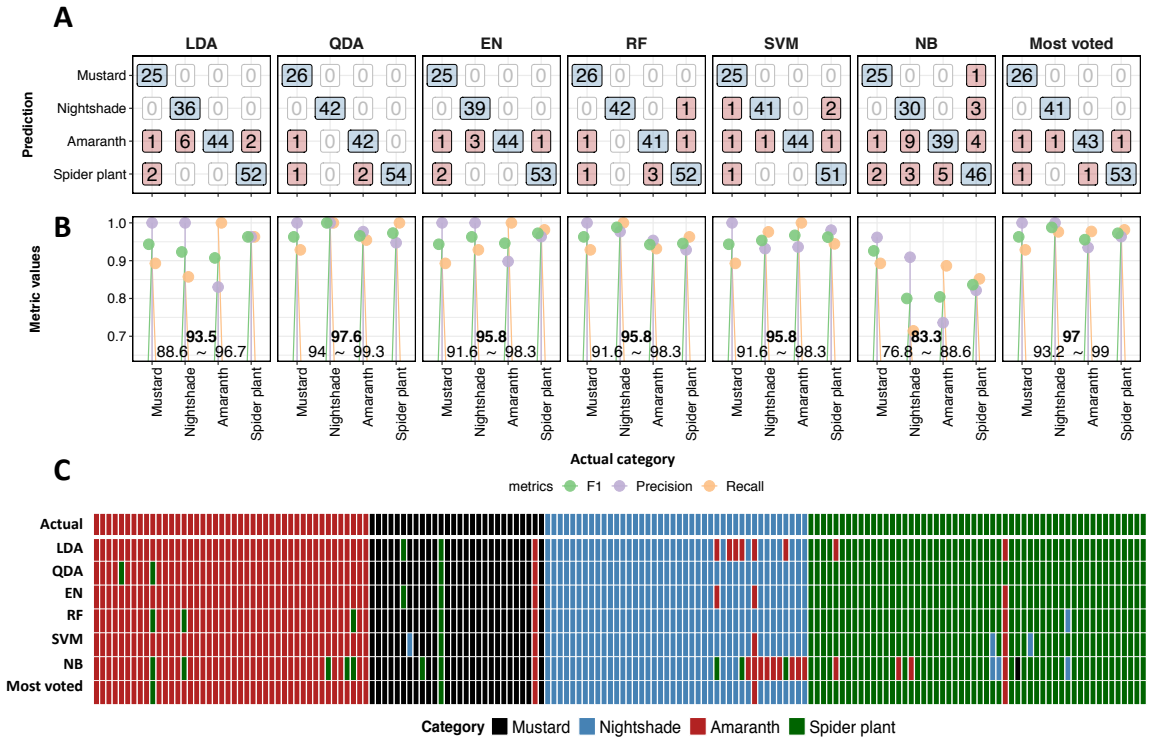


Figure IV-8. Prediction of African indigenous vegetables (AIVs) classification based on free amino acids (AAs) profile using machine learning (ML) techniques. The AIVs include four categories, i.e., Ethiopian mustards, African nightshades, amaranths and spider plants. The ML methods include linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), elastic net-regularized regression (EN), random forest (RF), support vector machine (SVM) and naïve Bayes (NB), as well as a simple ensembled method counting the most voted category of all prior models. Models are trained and tested using 70/30 split using category-site stratified sampling. (**A**), confusion matrix. (**B**), the prediction metrics, with precision, recall and F1 values for prediction of each category, with the model's overall accuracy shown in bold followed by 95% confidence interval. Both (**A**) and (**B**) apply the same abscissa. Note that the precision in the context of machine learning is different from the one in the context of LC-MS method validation. (**C**), prediction heatmap showing sample-wise results of each model compared with the actual category.

4 Conclusion

The HILIC UHPLC-MS/MS method developed and validated in this work allowed for confident analysis of underivatized AAs. Superior to literature of similar work, this method featured improved sensitivity, robust linear range and higher throughput, as well as a simpler and cleaner experimental set up. As two most critical aspects of method development, the removal of mineral buffer from the mobile phase was key for sensitivity boost without compromising the chromatographic performance. Secondly, addition of hydrochloric acid to the sample preparation solvent was of vital importance to maintain calibration linearity of the basic AAs. Using the developed methods, the free AAs in a total of 544 differently sourced AIVs were successfully determined. The four categories of AIVs including African nightshades, amaranth, spider plant and Ethiopian mustard were predicted based on the free AA profile using various ML methods with satisfactory accuracy. In addition, an R Shiny based online interactive ML platform was established, allowing for ML models train-test simulation, and rapid prediction of unknown AIV samples using the analyzed free AAs profile.

The future work involves analysis of more metabolites beyond the analyzed 21 AAs in the AIVs to provide more statistical power for classification prediction. Additional ML methods such as K-nearest neighbors and neural network, etc., will be applied for performance comparison. A more sophisticated ensemble method needs to be developed to take full advantage of the different strengths of varied ML models for prediction accuracy improvement. ML model interpretation as an emerging technique will be applied to render model readability and transparency [49]. Besides, based on the high content of free AAs in AIVs, it is of keen interest to further analyze the total protein content and

quality to shed light on the nutrition value. An online platform will be established for automatic computation of complete protein formula based on AIVs and other common foodstuff.

5 References

[1] Herforth AW. Promotion of traditional African vegetables in Kenya and Tanzania: a case study of an intervention representing emerging imperatives in global nutrition. Cornell University, 2010.

[2] Yang RY, Ojiewo C. African Nightshades and African Eggplants: Taxonomy, Crop Management, Utilization, and Phytonutrients. African Natural Plant Products Volume II: Discoveries and Challenges in Chemistry, Health, and Nutrition. ACS Symposium Series2013, p. 137-165.

[3] Neugart S, Baldermann S, Ngwene B, Wesonga J, Schreiner M. Indigenous leafy vegetables of Eastern Africa - A source of extraordinary secondary plant metabolites. *Food Res Int* (all book title or journal should be either in the format of Italic or underline) 2017;100:411-422.

[4] Jimenez-Aguilar DM, Grusak MA. Evaluation of Minerals, Phytochemical Compounds and Antioxidant Activity of Mexican, Central American, and African Green Leafy Vegetables. Plant Foods Hum Nutr 2015;70:357-64.

[5] Uusiku NP, Oelofse A, Duodu KG, Bester MJ, Faber M. Nutritional value of leafy vegetables of sub-Saharan Africa and their potential contribution to human health: A review. J Food Compost Anal 2010;23:499-509.

[6] Kamga RT, Kouamé C, Atangana AR, Chagomoka T, Ndango R. Nutritional Evaluation of Five African Indigenous Vegetables. Journal of Horticultural Research 2013;21:99-106.

[7] Rigas PG. Review: Liquid Chromatography—Post-Column Derivatization for Amino Acid Analysis: Strategies, Instrumentation, and Applications. Instrumentation Science & Technology 2012;40:161-193.

[8] Moor S. Chromatographic determination of amino acid by the use of automatic recording equipment. Methods and Enzymology 1963;6:819-831.

[9] Piraud M, Vianey-Saban C, Petritis K, Elfakir C, Steghens JP, Bouchu D. Ion-pairing reversed-phase liquid chromatography/electrospray ionization mass spectrometric analysis of 76 underivatized amino acids of biological interest: a new tool for the diagnosis of inherited disorders of amino acid metabolism. Rapid Commun Mass Spectrom 2005;19:1587-602.

[10] Waterval WA, Scheijen JL, Ortmans-Ploemen MM, Habets-van der Poel CD, Bierau J. Quantitative UPLC-MS/MS analysis of underivatised amino acids in body fluids is a reliable tool for the diagnosis and follow-up of patients with inborn errors of metabolism. Clin Chim Acta 2009;407:36-42.

[11] Fürst P, Pollack L, Graser T, Godel H, Stehle P. Appraisal of four pre-column derivatization methods for the high-performance liquid chromatographic determination of free amino acids in biological materials. Journal of Chromatography A 1990;499:557-569.

[12] Zheng N, Xiao H, Zhang Z, Gao X, Zhao J. Rapid and sensitive method for determining free amino acids in plant tissue by high-performance liquid chromatography with fluorescence detection. Acta Geochimica 2017;36:680-696.

[13] Jr JWH, Brooks A. Improved Amino Acid Methods using Agilent ZORBAX Eclipse Plus C18 Columns for a Variety of Agilent LC Instrumentation and Separation Goals. Agilent Technologies Application Note 2010:Publication Part Number: 5990-4547EN. https://www.agilent.com/search/?Ntt=Improved%20Amino%20Acid%20Methods%20using%20Agilent%20ZORBAX%20Eclipse%20Plus%20C18%20Columns%20for%20a%20Variety%20of%20Agilent%20LC%20Instrumentation%20and%20Separation%20Goals.

[14] Salazar C, Armenta JM, Shulaev V. An UPLC-ESI-MS/MS Assay Using 6-Aminoquinolyl-N-Hydroxysuccinimidyl Carbamate Derivatization for Targeted Amino Acid Analysis: Application to Screening of Arabidopsis thaliana Mutants. Metabolites 2012;2:398-428.

[15] Boogers I, Plugge W, Stokkermans YQ, Duchateau AL. Ultra-performance liquid chromatographic analysis of amino acids in protein hydrolysates using an automated pre-column derivatisation method. J Chromatogr A 2008;1189:406-9.

[16] Fiechter G, Pavelescu D, Mayer HK. UPLC analysis of free amino acids in wines: profiling of on-lees aged wines. J Chromatogr B Analyt Technol Biomed Life Sci 2011;879:1361-6.

[17] Unwin RD, Griffiths JR, Whetton AD. Simultaneous analysis of relative protein expression levels across multiple samples using iTRAQ isobaric tags with 2D nano LC-MS/MS. Nat Protoc 2010;5:1574-82.

[18] Kaspar H, Dettmer K, Chan Q, Daniels S, Nimkar S, Daviglus ML, Stamler J, Elliott P, Oefner PJ. Urinary amino acid analysis: a comparison of iTRAQ-LC-MS/MS, GC-MS, and amino acid analyzer. J Chromatogr B Analyt Technol Biomed Life Sci 2009;877:1838-46.

[19] Held PK, White L, Pasquali M. Quantitative urine amino acid analysis using liquid chromatography tandem mass spectrometry and aTRAQ reagents. J Chromatogr B Analyt Technol Biomed Life Sci 2011;879:2695-703.

[20] Shimbo K, Yahashi A, Hirayama K, Nakazawa M, Miyano H. Multifunctional and highly sensitive precolumn reagents for amino acids in liquid chromatography/tandem mass spectrometry. Anal Chem 2009;81:5172-9.

[21] Smon A, Cuk V, Brecelj J, Murko S, Groselj U, Zerjav Tansek M, Battelino T, Repic Lampret B. Comparison of liquid chromatography with tandem mass spectrometry and ion-exchange chromatography by post-column ninhydrin derivatization for amino acid monitoring. Clin Chim Acta 2019;495:446-450.

[22] Konn C, Magnér J, Charlou J-L, Holm NG, Alsberg T. A Method for Detection of Trace Concentrations of Underivatized Amino Acid in Hydrothermal Fluids by Ion-Pairing Reversed-Phase UPLC-ESI-QTOF-MS. American Journal of Analytical Chemistry 2015;06:313-324.

[23] McCalley DV. Is hydrophilic interaction chromatography with silica columns a viable alternative to reversed-phase liquid chromatography for the analysis of ionisable compounds? J Chromatogr A 2007;1171:46-55.

[24] Jandera P, Janas P. Recent advances in stationary phases and understanding of retention in hydrophilic interaction chromatography. A review. Anal Chim Acta 2017;967:12-32.

[25] Gokmen V, Serpen A, Mogol BA. Rapid determination of amino acids in foods by hydrophilic interaction liquid chromatography coupled to high-resolution mass spectrometry. Anal Bioanal Chem 2012;403:2915-22.

[26] Guo S, Duan JA, Qian D, Tang Y, Qian Y, Wu D, Su S, Shang E. Rapid determination of amino acids in fruits of Ziziphus jujuba by hydrophilic interaction ultra-high-performance liquid chromatography coupled with triple-quadrupole mass spectrometry. J Agric Food Chem 2013;61:2709-19.

[27] Gao J, Helmus R, Cerli C, Jansen B, Wang X, Kalbitz K. Robust analysis of underivatized free amino acids in soil by hydrophilic interaction liquid chromatography coupled with electrospray tandem mass spectrometry. J Chromatogr A 2016;1449:78-88.

[28] Prinsen H, Schiebergen-Bronkhorst BGM, Roeleveld MW, Jans JJM, de Sain-van der Velden MGM, Visser G, van Hasselt PM, Verhoeven-Duif NM. Rapid quantification of underivatized amino acids in plasma by hydrophilic interaction liquid chromatography (HILIC) coupled with tandem mass-spectrometry. J Inherit Metab Dis 2016;39:651-660.

[29] Tsochatzis ED, Begou O, Gika HG, Karayannakidis PD, Kalogiannis S. A hydrophilic interaction chromatography-tandem mass spectrometry method for amino acid profiling in mussels. J Chromatogr B Analyt Technol Biomed Life Sci 2017;1047:197-206.

[30] Zhang X, Chen H, Wu D, Gu W, Sun X, Chen J, Wu Q. Determination of Free Amino Acids in Three Species of Duckweed (Lemnaceae). J Food Qual 2018;2018:1-15.

[31] Xu YF, Lu W, Rabinowitz JD. Avoiding misannotation of in-source fragmentation products as cellular metabolites in liquid chromatography-mass spectrometry-based metabolomics. Anal Chem 2015;87:2273-81.

[32] Byrnes DR, Dinssa FF, Weller SC, Simon JE. Elemental Micronutrient Content and Horticultural Performance of Various Vegetable Amaranth Genotypes. J Am Soc Hort Sci 2017;142:265-271.

[33] Yuan B, Byrnes D, Giurleo D, Villani T, Simon JE, Wu Q. Rapid screening of toxic glycoalkaloids and micronutrients in edible nightshades (*Solanum* spp.). J Food Drug Anal 2018;26:751-760.

[34] Dinssa FF, Hanson P, Ledesma DR, Minja R, Mbwambo O, Tilya MS, Stoilova T. Yield of Vegetable Amaranth in Diverse Tanzanian Production Environments. HortTechnology 2019;29:516-527.

[35] Yuan B, Dinssa FF, Simon JE, Wu Q. Simultaneous quantification of polyphenols, glycoalkaloids and saponins in African nightshade leaves using ultra-high-performance liquid chromatography-tandem mass spectrometry with acid assisted hydrolysis and multivariate analysis. Food Chem 2019:126030.

[36] Yuan B, Zhao D, Du R, Kshatriya D, Bello NT, Simon JE, Wu Q. A highly sensitive ultra-high performance liquid chromatography/tandem mass spectrometry method with in-source fragmentation for rapid quantification of raspberry ketone. J Food Drug Anal 2019;27:778-785.

[37] Administration UFaD. Guidance for Industry: Bioanalytical Method Validation. 2018.

[38] Kruve A, Rebane R, Kipper K, Oldekop M-L, Evard H, Herodes K, Ravio P, Leito I. Tutorial review on validation of liquid chromatography–mass spectrometry methods: Part I. Anal Chim Acta 2015;870:29-44.

[39] Kruve A, Rebane R, Kipper K, Oldekop ML, Evard H, Herodes K, Ravio P, Leito I. Tutorial review on validation of liquid chromatography-mass spectrometry methods: part II. Anal Chim Acta 2015;870:8-28.

[40] Johnson RA, Wichern DW. Applied multivariate statistical analysis. 6th edition: Pearson India; 2015.

[41] Friedman J, Hastie T, Tibshirani R. The elements of statistical learning: Springer series in statistics New York; 2001.

[42] R-Core-Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/. 2013.

[43] Grolemund G, Wickham H. R for data science. 2018.

[44] Grumbach ES, Fountain KJ. Comprehensive Guide to HILIC: Hydrophilic Interaction Chromatography: Waters; 2010.

[45] Snyder LR, Kirkland JJ, Dolan JW. Introduction to modern liquid chromatography: John Wiley & Sons; 2011.

[46] Ito H, Ueno H, Kikuzaki H. Construction of a free-form amino acid database for vegetables and mushrooms. Integr Food Nutr Metab 2017;4:1-9.

[47] Kim YB, Uddin MR, Lee M-K, Kim S-J, Kim HH, Chung E, Lee J-H, Park SU. Accumulation of Free Amino Acids in Different Organs of Green and Red Mustard Cultivars. Asian J Chem 2014;26:396-398.

[48] Oliveira AP, Pereira DM, Andrade PB, Valentao P, Sousa C, Pereira JA, Bento A, Rodrigues MA, Seabra RM, Silva BM. Free amino acids of tronchuda cabbage (Brassica oleracea L. Var. costata DC): influence of leaf position (internal or external) and collection time. J Agric Food Chem 2008;56:5216-21.

[49] Molnar C. A guide for making black box models explainable. URL: https://christophm github io/interpretable-ml-book/(дата обращения: 2803 2019) 2018.

CHAPTER V. OTHER LEAFY MICRONUTRIENTS

1 Introduction

This chapter is devoted to the quantitative determination of vitamin-A precursor β-carotene, as well as other nutrients including vitamin E in the leafy aerial parts of African nightshades. In addition, determination of total polyphenol (TPP) content using Folin-Ciocalteu's assay and antioxidant activity using ABTS radical was also performed. The TPP reflects the total polyphenol level using gallic acid equivalent, and may be viewed as supplementary analysis of prior quantification of individual polyphenol compounds discussed in CHAPTER III. In addition, TPP level also reflects the content of other reducing compounds, such as ascorbic acid, etc. The total antioxidant assay also reflects the content of polyphenols, ascorbic acid as well as other reducing compounds.

2 Materials and Methods

2.1 Chemical reagents

Standards β-carotene and vitamin E (α-tocopherol), Folin Ciocalteu's phenol reagent, 6-hydroxy-2,5,7,8-tetramethylchromane-2-carboxylic acid (Trolox) and 2,2'-azino-bis (3-ethylbenzothiazoline-6-sulphonic acid (ABTS) from Sigma-Aldrich (St. Louis, MO, USA). Gallic acid was purchased from Acros Organics (Belgium, WI, USA) and acetone from BDH Chemicals (Radnor, PA, USA). Methanol, ethyl acetate, *tert*-butyl methyl ether were purchased from Fisher Scientific (Fair Lawn, NJ, USA).

2.2 Equipment

Waters 2695 HPLC (Milford, MA, USA) was used for β-carotene and vitamin E measurement, which was equipped with a quaternary pump and a diode array detector. The separation was achieved by YMC-C30 carotenoid C30 column, 5 μm, 250 × 4.6 mm (YMC Co., Ltd). The software was Millennium 4.00. Bio-Tek Synergy HT Multi-Mode Microplate reader (Winooski, VT, USA) was used for spectrophotometric measurement for total polyphenol assay and antioxidant assay. The software used was Bio-Tek KC4 Version 3.4.

2.3 . Sample preparation

The plant materials were from the same batch as in CHAPTER II. The plant identity is also presented in **Table V-1**. For the analysis of β-carotene and vitamin E, about 500 mg sample was weighted accurately and extracted by 5 mL acetone in a two dram vial, fully vortexed and sonicated for 30 s. The extract was then centrifuged at 3000 rpm for 5 min with the supernatant decanted to an 8 dram amber vial. The sample was subsequently extracted in like manner by another 5 mL acetone and 2 mL *tert*-butyl methyl ether, respectively, with the supernatants combined in the 8 dram vial. The supernatant was then rotovapped to yield the dry extract. The dried residue was then reconstituted into 2 mL 1:1 ethyl acetate: methanol, followed by centrifugation at 3000 rpm for 10 min, and the supernatant was ready for HPLC analysis. The extract for each entry was prepared in triplicate.

For TPP assay and antioxidant assay, around 200 mg of the sample was accurately weighed and extracted with 25 mL 70 % methanol with 0.1 % formic acid. Each extract was fully vortexed, sonicated in a water bath for 5 min, and then let stand still overnight

under room temperature [1-3]. The extract was centrifuged at 3000 rpm/min for 10 min. For TPP assay, the supernatant was directly ready for analysis. For antioxidant activity assessment, the supernatant was diluted twice by 70 % methanol with 0.1 % formic acid before the assay.

2.4 β-carotene and vitamin E analysis

The β-carotene and vitamin E content was analyzed using Waters HPLC. Mobile phase A was micron-filtered 98:2 methanol:1 M ammonium acetate, and mobile phase B was HPLC grade ethyl acetate, and both were manually degassed by sonication under vacuum. The gradient started from 0% to 35 % B in 0 to 8 min, 35% to 40 % in 8 to 28 min, 40% to 50 % in 28 to 32 min, and 50 % to 60 % in 32 to 36 min and held until 40 min. The injection volume was 10 µl. UV chromatograms for β-carotene and vitamin E were generated at their maximum peak absorption wavelength of 452 nm and 295 nm, respectively, and both compounds were identified in comparison with their retention time and UV-Vis spectrum of the authenticated standards. For the calibration curve, around 20 mg vitamin E and 10 mg β-carotene standards were accurately weighed and dissolved in 25 mL ethyl acetate as stock solutions, respectively [4]. The stock solutions were then diluted by ethyl acetate into series to generate work solutions of ~4 to ~150 µg/mL for both compounds. The β-carotene and vitamin E content in samples was reported as µg/g dry mass (DW).

2.5 . Total polyphenol (TPP)

The method of TPP assay was based on that proposed by Singleton [5] and Kao et al. [6]with modification. First, 50 mL Folin Ciocalteu's reagent was diluted by distilled water to 500 mL, then 900 μL Folin reagent was mixed with 80 μL sample extract followed by the addition of 400 μL saturated sodium carbonate solution. The reaction system was fully vortexed and let stand still for 1 h. Then 200 μL supernatant was transferred to a 96 multiple well and subject to absorption measurement under 765 nm. As to the calibration curve, 20.0 mg gallic acid was dissolved in 5 mL 70 % methanol as the stock solution, and diluted to a series of work solutions with concentrations ranging from 15.6 μg/mL to 250 μg/mL. 200 μL 70 % methanol was used as negative control. The assay for each calibrator and sample was triplicated. Total polyphenol content in samples was expressed as the amount of gallic acid equivalent (GAE) /g DW.

2.6 Antioxidant activity

The antioxidant activity assay was based on the method proposed by Re et al. [7] and Nagulsamy [8] with modification. First, 31.7 mg ABTS and 8.6 mg potassium persulfate were dissolved in 10 mL water and let stand still in darkness under room temperature for 12~16 h to form stable radical, and diluted to an absorption of ~1.3 at 734 nm. Next, 200 μl ABTS was mixed with 20 μl sample extract and let stand still for 15 min under room temperature, and the absorption was measured at 734 nm. For the calibration curve, 12.5 mg standard Trolox was dissolved in 5 mL pure ethanol as the stock solution, and diluted to series of work solutions with concentrations ranging from 19.5 μg/mL to 195.3 μg/mL. The assay for each calibrator and sample was conducted in triplicate. Antioxidant activity was expressed as the amount of Trolox equivalent (TE) /g DW.

2.7 Statistics

Data were represented in the form of mean ± the standard deviation (STD) of three replicates of each sample. Data was analyzed using student's *t*-test and one-way analysis of variance (ANOVA) followed by Tukey's Honest Significant Difference (HSD) test ($p < 0.05$). The relationship between TPP content and antioxidant activity was evaluated by Pearson's correlation. All statistics was performed using SAS University Edition (SAS Institute Inc., Cary, NC, USA).

3 Resutls and Discussion

3.1 β-carotene and vitamin E analysis

β-Carotene and vitamin E in nightshade extracts were identified and quantified against standards using UV detection at 454 nm and 295 nm, respectively. Representative chromatograms are shown in **Figure V-1**. Quantification results were summarized in **Table V-1**. β-Carotene in all four species was higher than the average level of a wide scope of household fresh vegetables reported by USDA National Nutrient Database (**Figure V-2**). A total of 10 entries were identified to be "a source" of vitamin A based on retinol equivalent (RE) according to standards published by Codex Alimentarius [9, 10]. The averaged β-carotene content across the five accessions of *S. nigrum* was 81.1 ± 40.7 μg/g DW. Except for *S. nigrum* PI 306400, all the other accessions contained higher level of β-carotene than that reported 4.66 ± 0.02 mg/100g DW or 46.6 ± 0.2 μg/g DW in a nutritive study of *S. nigrum* L. *var. virginicum* of Nigeria [11]. Up to 141.7 ± 11.7 μg/g DW, *S. nigrum* PI 312110 possessed the highest content in all accessions of *S. nigrum* and the other three species. In *S. scabrum*, SS 49 was found to contain the most abundant source of β-

carotene with up to 96.0 ± 2.8 μg/g DW. The mean level in all eight accessions of *S. scabrum* was 79.4 ± 15.2 μg/g DW, nearly twice the level reported by Mibei at el. at 46 ± 2.5 μg/g DW [12]. The β-carotene level in the single accession of *S. americanum* was 95.8 ± 3.8 μg/g DW, less than a fifth of the reported 52.1 ± 3.6 mg/100g DW in the literature [13]. In contrast, the β-carotene level of *S. villosum* was found to be high at 138.1 ± 4.0 μg/g DW and this is the report of β-carotene from this species.

Vitamin E content in all four species, except in a few entries, was among the top 25 % when compared with the USDA National Nutrient Database (**Figure V-2**). In the five accessions of *S. nigrum*, vitamin E content reached 150.3 ± 59.9 μg/g DW on average. The accession *S. nigrum* PI 312110 was found not only to be high in β-carotene, but also the richest source of vitamin E among all accessions of the four studied species at 229.7 ± 19.9 μg/g DW. In the eight entries of *S. scabrum*, the average level was 126.2 ± 44.3 μg/g DW, with BG 16 having the highest concentration of vitamin. The content in the one accession of *S. americanum* was 145.5 ± 5.2 μg/g, and in the single accession of *S. villosum* was 114.3 ± 5.0 μg/g.

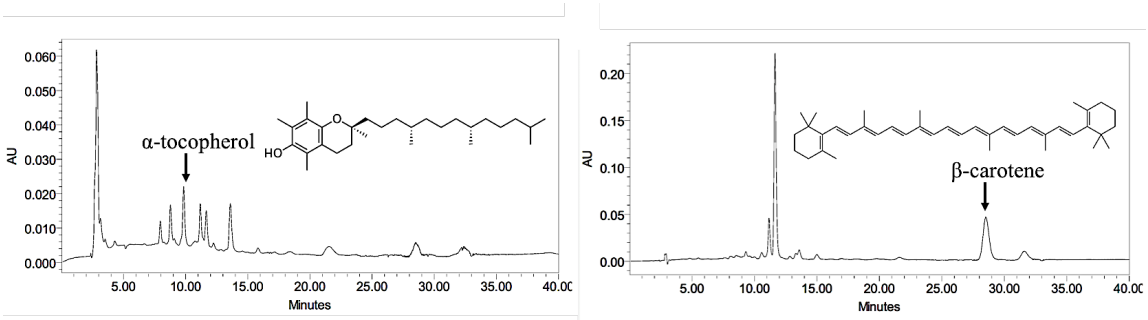


Figure V-1. UV/vis chromatogram of α-tocopherol at 295 nm and β-carotene at 454 nm from *S. nigrum* USDA Grif 14198.

Table V-1. Identification and nutrients content in 15 accessions of edible nightshades (*Solanum* spp.).

Sample Code †	Source *	Source ID	Vitamin E (μg/g DW)	β-carotene (μg/g DW)	ABTS (TE mg/g DW)	TPP (GAE mg/g DW)
S.n 1	Simlaw Seeds (Kenya)	--	106.9 ± 2.0^{def}	*78.4 ± 4.5^{bcd}*	16.80 ± 0.64^{gh}	8.54 ± 0.33^{gef}
S.n 2	USDA	PI 306400	92.0 ± 2.6^{gf}	28.1 ± 1.1^{f}	15.49 ± 0.69^{h}	7.25 ± 0.36^{h}
S.n 3	USDA	PI 312110	229.7 ± 19.9^{a}	*141.7 ± 11.7^{a}*	23.45 ± 0.68^{abc}	11.92 ± 0.49^{ab}
S.n 4	USDA	PI 381289	125.9 ± 7.6^{def}	70.6 ± 2.5^{ced}	18.34 ± 0.77^{gf}	9.58 ± 0.33^{de}
S.n 5	USDA	PI 381290	197.1 ± 17.9^{ab}	*86.7 ± 8.9^{bc}*	23.93 ± 1.12^{ab}	13.07 ± 0.42^{a}
S.s 1	AVRDC	SS 52	114.9 ± 8.6^{def}	*87.4 ± 3.7^{bc}*	21.36 ± 0.50^{dc}	10.8 ± 0.35^{bc}
S.s 2	AVRDC	Ex Hai	90.4 ± 3.7^{gf}	64.8 ± 0.3^{ed}	17.92 ± 0.45^{gf}	7.98 ± 0.26^{ghf}
S.s 3	AVRDC	SS 49; Olevolosi	141.3 ± 1.9^{de}	*96.0 ± 2.8^{b}*	22.46 ± 1.11^{bdc}	12.21 ± 0.53^{a}
S.s 4	AVRDC	SS 04.2	64.1 ± 2.8^{g}	65.2 ± 2.0^{ed}	18.11 ± 0.62^{gf}	9.39 ± 0.28^{de}
S.s 5	AVRDC	BG 16; Nduruma	192.5 ± 28.5^{ab}	*91.5 ± 13.0^{bc}*	25.00 ± 0.70^{a}	12.47 ± 0.42^{a}
S.s 6	AVRDC	BG-29	102.1 ± 11.7^{gef}	55.1 ± 7.3^{e}	19.14 ± 0.33^{ef}	9.13 ± 0.24^{def}
S.s 7	USDA	Grif 14198	121.1 ± 8.6^{def}	*87.7 ± 4.7^{bc}*	21.26 ± 0.59^{d}	10.98 ± 0.40^{bc}
S.s 8	USDA	PI 643126	183.4 ± 30.1^{bc}	*87.2 ± 14.0^{bc}*	16.22 ± 0.33^{gh}	7.75 ± 0.52^{gh}
S.a 1	USDA	PI 268152	145.5 ± 5.2^{dc}	*95.8 ± 3.8^{b}*	24.81 ± 0.09^{a}	12.44 ± 0.43^{a}
S.v 1	USDA	Grif 16939	114.3 ± 5.0^{def}	*138.1 ± 4.0^{a}*	20.99 ± 0.99^{ed}	10.26 ± 0.40^{dc}

The results were reported as mean ± standard deviation (n = 3). $^{a\text{-}h}$, values with same superscript letters in the same column are not significantly different (ANOVA with Tukey's HSD test, $p < 0.05$). †. S.n: *Solanum nigrum*; S.s: *S. scabrum*; S.a: *S. americanum*; S.v: *S. villosum*. ND, not detected. *, Seeds from USDA were maintained by Plant Genetic Resources Conservation Unit, Griffin, GA. Seeds from AVRDC were maintained by the regional center of east and southern Africa, Arusha, Tanzania. The plant samples studied in this chapter was from the same batch as in CHAPTER II. Italicized values of β-carotene contents were identified to be "a source" of vitamin A based on retinol equivalent (RE), assuming 90 % moisture content in fresh vegetables [9, 10].

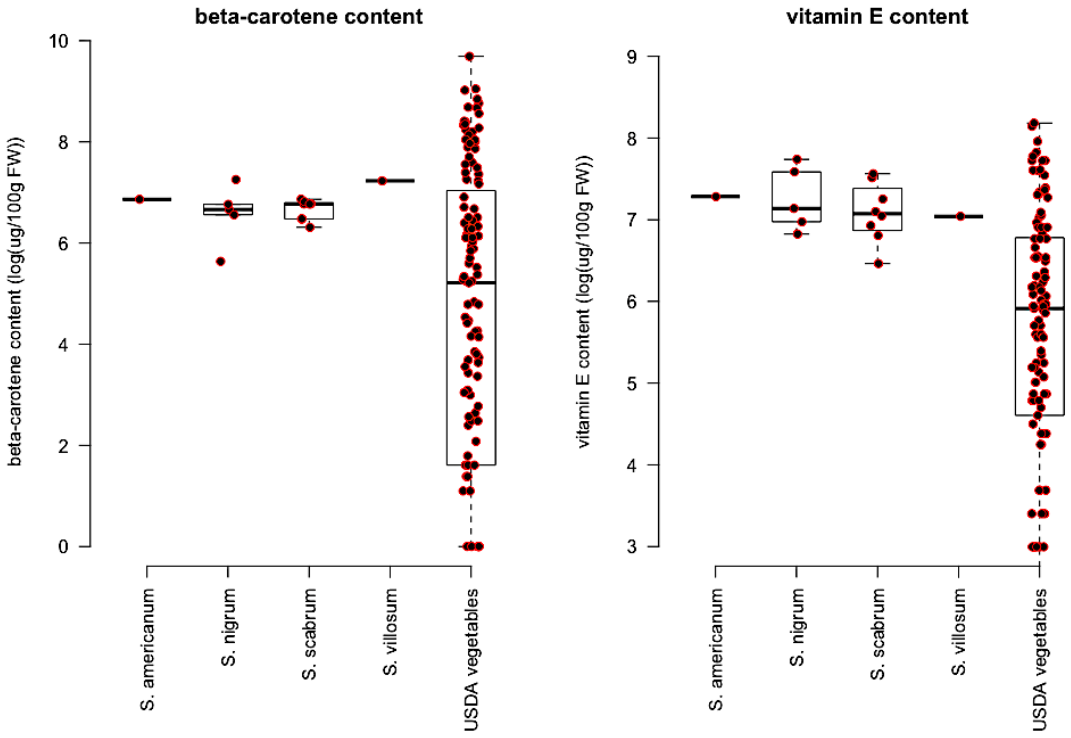


Figure V-2. Content of beta-carotene and vitamin E in the edible nightshade species compared with those in the diverse vegetables species reported in USDA website (*https://fdc.nal.usda.gov/*). Concentration was expressed in natural logarithm due to the large dynamic range of production of β-carotene and vitamin E. 90 % moisture content was assumed in calculation of β-carotene and vitamin E in fresh edible nightshades.

Data Source: USDA National Nutrient Database for Standard Reference 28 Software v.3.7.1 2017-03-29

Nutrients: Vitamin E (alpha-tocopherol) (mg); Carotene, beta (*μ*g)

Food Groups: Vegetables and Vegetable Products

Food Subset: All Foods, *with "raw" items manually selected by the author for data comparability.*

Ordered by: Food Name

Measured by:100 g

Report Run at: 06-12-2017T17:53

3.2 Total polyphenol content and antioxidant activity

The TPP level and antioxidant activity values were summarized in **Table V-1.** The TPP level in the five accessions of *S. nigrum* averaged 10.07 ± 2.40 mg GAE/g DW, and in the eight accessions of *S. scabrum* was 10.09 ± 1.80 mg GAE/g DW. *Solanum scabrum* PI 643126 and Grif 14198 was reported in another study to contain free phenolic acids at a level of 12.83 ± 0.41 mg GAE/g DW and 15.65 ±1.08 mg GAE/g DW, respectively, which was correspondingly 65% and 42.5% higher than the TPP level found in this study [14].

Antioxidant activity in the five accessions of *S. nigrum* was averaged to be 19.6 ±3.9 mg TE/g DW, and in the eight accessions of *S. scabrum* at 20.2 ± 2.8 mg TE/g DW. In comparison, Jimenez-Aguilar using AAPH radical reported the antioxidant activity of *S. scabrum* PI 643126 and Grif 14198 to be 34.5 ± 1.95 μmol TE/g FW and 36.17± 1.47 μmol TE/g FW, respectively, or 50.15 ± 2.83 mg/g DW and 51.35 ± 2.09 mg/g DW in accordance [14]. Among samples in this study, there was a close correlation between TPP content and antioxidant capability with R^2=0.95 (n=15), suggesting that either the TPP content or ABTS antioxidant activity may be used for quality control purpose.

4 References

[1] Zhen J, Villani TS, Guo Y, Qi Y, Chin K, Pan MH, Ho CT, Simon JE, Wu Q. Phytochemistry, antioxidant capacity, total phenolic content and anti-inflammatory activity of Hibiscus sabdariffa leaves. Food Chem 2016;190:673-80.

[2] Bai Y. Simultaneous quantification of four saponins, three alkaloids and three fatty acids in *Solanum nigrum* Linn. by HPLC-ELSD. J Med Plants Res 2012;6.

[3] Do QD, Angkawijaya AE, Tran-Nguyen PL, Huynh LH, Soetaredjo FE, Ismadji S, Ju YH. Effect of extraction solvent on total phenol content, total flavonoid content, and antioxidant activity of Limnophila aromatica. J Food Drug Anal 2014;22:296-302.

[4] Denardin CC, Hirsch GE, da Rocha RF, Vizzotto M, Henriques AT, Moreira JCF, Guma F, Emanuelli T. Antioxidant capacity and bioactive compounds of four Brazilian native fruits. J Food Drug Anal 2015;23:387-398.

[5] Singleton V, Rossi JA. Colorimetry of total phenolics with phosphomolybdic-phosphotungstic acid reagents. Am J Enol Vitic 1965;16:144-158.

[6] Kao F-J, Chiu Y-S, Chiang W-D. Effect of water cooking on antioxidant capacity of carotenoid-rich vegetables in Taiwan. J Food Drug Anal 2014;22:202-209.

[7] Re R, Pellegrini N, Proteggente A, Pannala A, Yang M, Rice-Evans C. Antioxidant activity applying an improved ABTS radical cation decolorization assay. Free Radic Biol Med 1999;26:1231-7.

[8] Nagulsamy P, Ponnusamy R, Thangaraj P. Evaluation of antioxidant, anti-inflammatory, and antiulcer properties of Vaccinium leschenaultii Wight: A therapeutic supplement. J Food Drug Anal 2015;23:376-386.

[9] Food and Agriculture Organization of the United Nations WHO. GUIDELINES FOR USE OF NUTRITION AND HEALTH CLAIMS (CAC/GL 23-1997). Codex Alimentarius.

[10] Food and Agriculture Organization of the United Nations WHO. GUIDELINES ON NUTRITION LABELLING (CAC/GL 2-1985). Codex Alimentarius.

[11] Akubugwo I, Obasi A, Ginika S. Nutritional potential of the leaves and seeds of black nightshade-Solanum nigrum L. Var virginicum from Afikpo-Nigeria. Pak J Nutr 2007;6:323-326.

[12] Mibei EK, Ojijo NKO, Karanja SM, Kinyua JK. Compositional attributes of the leaves of some indigenous African leafy vegetables commonly consumed in Kenya. Ann: Food Sci Technol 2011;12:146-154.

[13] Serrano J, Goni I, Saura-Calixto F. Determination of beta-carotene and lutein available from green leafy vegetables by an in vitro digestion and colonic fermentation method. J Agric Food Chem 2005;53:2936-40.

[14] Jimenez-Aguilar DM, Grusak MA. Evaluation of Minerals, Phytochemical Compounds and Antioxidant Activity of Mexican, Central American, and African Green Leafy Vegetables. Plant Foods Hum Nutr 2015;70:357-64.

CHAPTER VI. BERRY PHYTOCHEMICAL IDENTIFICATION

1 Introduction

While the leaves of African nightshades are popularly consumed as green vegetables, the blueberry-like berries despite their prolific production, however, are generally discarded without consumption possibly due to perception of toxicity [1]. Unrecognition of berries in this area is in contrast to some cultural practices in Europe where berries are considered edible, consumed either raw or processed as food ingredients [2]. Such discrepancy may result from variation of berries in toxicity and nutrition level due to genetic variance or environmental difference, or simply a misconception of berry edibility, and has not been brought under thorough investigation. This remains a concerning issue as on one hand, the true consumption safety of berries remains unexamined and unaware consumption of toxic berries could cause long-term or even acute health hazard, while on the other hand berries with their large production in sub-Saharan Africa remain an underutilized agricultural resource and, if indeed edible, an untapped new food supply.

Evaluation of berry toxicity and nutrition potential requires knowledge of the relevant phytochemical compositions, and such information so far remains incomplete. Identification of the phytochemicals has nowadays been greatly advanced by application of liquid chromatography hyphenated with mass spectrometry (LC-MS). Particularly, the enriched information acquired from mass spectra with known scaffold-fragmentation pattern allows for structural elucidation of compounds even novel ones in a plant extract

complex prior to the tedious labor of isolation and purification of each individual compound [3].

The aim of this study was to comprehensively identify the phytochemicals of *S. scabrum* berries using different techniques of LC/UV-vis/MS complemented with acid-assisted hydrolysis. This work would lay the foundation for quantitative comparison of berry phytochemical profile and re-evaluation of toxicity and nutrition. It would also facilitate exploration of new food supply, reduce malnutrition and create value addition for producers in sub-Saharan Africa.

2 Materials and Methods

2.1 Chemical reagents

Authentic reference standards neochlorogenic acid, chlorogenic acid, quercetin and dioscin were purchased from Sigma-Aldrich (St. Louis, MO), isorhamnetin from Carl Roth (Karlsruhe, Germany), solasodine from MP Biomedicals (Santa Ana, CA), solamargine from MedChem Express (Monmouth Junction, NJ), diosgenin, delphinidin chloride, petunidin chloride and malvidin chloride from ChromaDex, Inc. (Irvine, CA), and tigogenin from ALB Technology Limited (Mongkok Kowloon, Hong Kong, China). Methanol, concentrated hydrochloric acid, trifluoroacetic acid, and HPLC grade water and acetonitrile modified with 0.1 % formic acid were purchased from Fisher Scientific (Fair Lawn, NJ).

2.2 Plant Materials

Seeds of eight entries of *S. scabrum* from World Vegetable Center (WorldVeg), U.S. Department of Agriculture (USDA, US Germplasm Repository System, Tifton, Georgia) and commercial sources were collected (Table S1), and were sown and germinated in 2016 at the New Jersey Agricultural Experiment Station (NJAES) Research Greenhouse of Rutgers University. After four-weeks, the seedlings were then transplanted to NJAES Horticultural Research Farm #3. Berries of different maturity were manually collected in September 2016, identified as immature, green in both skin and flesh with firm texture; half-mature, with purple skin and light-green flesh; and mature, deep purple throughout the entire berry, soft and juicy. The final field harvest of berries occurred after a frost in early November 2016.

2.3 Instrumentation and equipment.

An Agilent 1100 series LC/MSD instrument (Agilent Technologies, Waldbronn, Germany) was used for the phytochemical characterization. The HPLC was composed of independent modules including an auto-degasser, quaternary pump, thermostatted column compartment and a diode-array detector (DAD). Column Polaris 180Å Amide-C18, 250 x 4.6 mm, 3 μm (Santa Clara, CA) was used for phytochemical profiling by method (a) and (b), and column Phenomenex Luna C18 (2), 150 x 4.60, 5μm (Torrance, CA) was used for analysis of aglycones freed from corresponding glycosides after acid hydrolysis by method (c). The HPLC-MS interface was electrospray ionization source (ESI) and the MS had an ion trap analyzer. The software was Agilent ChemStation A.08.03, LC/MSD Trap Control 5.1, and DataAnalysis 2.2

2.4 Sample preparation.

The harvested berries were dehydrated in an air-circulated 40 °C oven to visual dryness and crispiness and then ground using a shearing-action mill with 2 mm mesh filter screen on the out-flow. Samples were stored in dual-layer re-sealable bags in a cool place out of direct light. For phytochemical profiling by HPLC-MS method (a) (***section 2.5***) and method (b) (***section 2.6***), ~ 200 mg samples were extracted by 25 mL 70 % methanol acidified with 0.1 % formic acid, vortexed and then sonicated for 20 min. Next, the extract was allowed to stand at room temperature overnight. The extract was then centrifuged at 12,000 rpm/min for 10 min, and the supernatant was separated for analysis. For aglycone identification by HPLC-MS method (c) (***section 2.7***), ~200 mg sample was hydrolyzed by 25 mL anhydrous 0.5 mol·L^{-1} HCl methanol and incubated under 70 °C for 90 min. The hydrolysate was then cooled down to room temperature, centrifuged and the supernatant was separated for analysis.

2.5 Identification of polyphenols, alkaloids and saponins by method (a).

The HPLC mobile phase consisted of water with 0.1 % formic acid (A) and acetonitrile with 0.1 % formic acid (B). The flow rate was 1 mL·min^{-1}. The gradient was 10 to 15 % B from 0 to 5 min; 15 to 30 % B from 5 to 35 min; 30 to 50 % B from 35 to 55 min; and 50 to 60 % B from 55 to 60 min. The column was equilibrated with 10 % B for 10 min between injections and thermostatted under 25 °C. The injection volume was 10 µl. The DAD was set at 210 nm, 254 nm, and 370 nm, with 400 nm as reference and band width of 10 nm. The scanning range was from 200 to 550 nm with 2.0 nm scanning step. A third of HPLC eluent after column was split into MS. With respect to ESI, the nebulizer

needle voltage was positive 3500 V. High purity nitrogen was used as nebulizing gas at 40 psi and drying gas at 350 °C with a flow rate of 10 $L \cdot min^{-1}$. High purity helium was used as the collision gas. The collision energy was set at low level of 80 % (MS-CID 80%, CID for collision induced dissociation) to characterize the sequence of glycosylation and acylation and identification of aglycone, and at moderately high level of 150 % (MS-CID 150 %) to generate additional aglycone fragments for structural elucidation in separate runs. The scanning mode was set at positive polarity with scanning range from 100-1500 *m/z*. Ion counts control (ICC) was set at targeted 40,000 with maximum accumulation time of 300 ms.

2.6 Identification of anthocyanins by method (b)

The experimental conditions remained the same as method (a) except for the following specification. The mobile phase A was water with 0.4 % trifluoroacetic acid (TFA), and mobile phase B was acetonitrile with 0.4 % TFA. The gradient was 13 to 17 % from 0 to 2 min; 17 to 20 % from 2 to 13 min; 20 to 30 % from 13 to 25 min; and 30 to 50 % from 25 to 45 min. The column thermostat was set at 30 °C. The DAD was set at 520 nm with scanning range from 200 to 600 nm. Only MS-CID 80% was used for compound fragmentation.

2.7 Identification of aglycones by method (c)

In order to facilitate aglycone identification by comparison with authentic standards and structural elucidation, and given the limited sensitivity and scanning speed of MS^n (n>3), plant samples were hydrolyzed to break down glycosidic bonds and aglycones thus

freed were chromatographed by HPLC-MS and MS/MS. The parameters remained the same as method (a) except the following specification. The gradient was 10 to 24 % B from 0 to 15 min, held isocratically with 24 % B from 15 to 22 min, then 24 to 30 % from 22 to 35 min, 30 to 40 % from 35 to 45 min, followed by rapid increase to 80 % at 50 min, then 80 to 90 % from 50 to 60 min, and isocratically held at 90 % until 70 min.

For aglycones whose structure remained ambiguous by HPLC-MS experiments, the structure was further analyzed by MS/MS (product ion scan) in time-sectioned manner or in separate runs. Protonated aglycones were selected as the precursor ion and fragments were scanned from 100 to 500 *m/z*. The collision energy was pre-optimized by syringe infusion method using representative aglycone standards for corresponding categories, i.e., fragmentation amplitude of 2.5 was selected for aglycones of flavonols by using quercetin standard, and amplitude of 7.0 for aglycones of glycoalkaloids by using solasodine standard.

2.8 Nomenclature.

The nomenclature of fragmentation pathway of flavonoids described in prior research [4] was used in this study. The labels ${}^{i,j}A^+$ and ${}^{i,j}B^+$ correspond to the fragmental ions containing A and B ring, respectively, formed by cleavage of the i and j bonds of C ring. Fragmentation pathway of alkaloids was named in similar manner, with broken bonds specified by hyphen-connected carbon numbers or heteroatom-symbols they connect. For instance, the label ${}^{17\text{-}20,22\text{-}O}ABCD^+$ indicated fragmental ions containing the A-B-C-D rings, formed by the cleavage of bond between C17 and C20 (noted as “17-22”), and a second

bond between C22 and the adjacent oxygen atom in ring E (noted as “22-O”, “O” for oxygen atom).

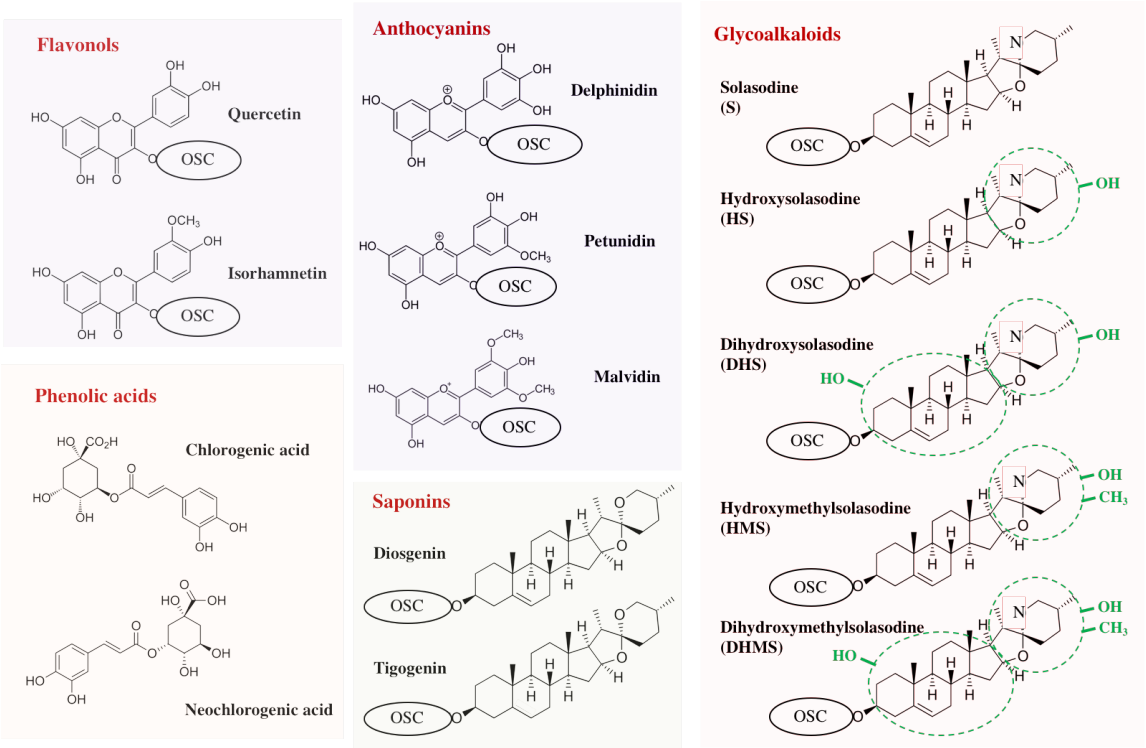


Figure VI-1. *Solanum scabrum* and *S. nigrum* berries glycoalkaloids and other major phytochemicals identified. The abbreviation of the glycosidic oligosaccharide chain is noted as OSC, that of the aglycone is noted in the parenthesis following the full spelled name. For glycoalkaloids, the hydroxymethylated derivatives of solasodine, including HMS and DHMS, may include corresponding methoxylated counterparts or other likely isomers including stereoisomers.

3 Results and Discussion

A total of 54 phytochemicals in mature berries from eight different genetic sources as well as in berries at different maturation stages were identified including phenolic acids, flavonols, anthocyanins, glycoalkaloids and saponins by analyzing their UV-Vis and MS data, and part of them were further confirmed by comparison with the corresponding

standards. An overview of representative molecular structures is shown in **Figure VI-1** (prior page). All compound identities with retention time, UV-vis spectrum, major fragmental ions using methods (a) and (b) are summarized in **Table VI-1**. Representative chromatograms and mass spectra acquired by method (a) are shown in **Figure VI-2** and **Figure A-12**, respectively, and those generated by method (b) are shown in **Figure VI-3** and **Figure A-13**, accordingly.. Chromatographed aglycones freed after acid assisted hydrolysis by method (c) was shown in **Figure VI-4**.

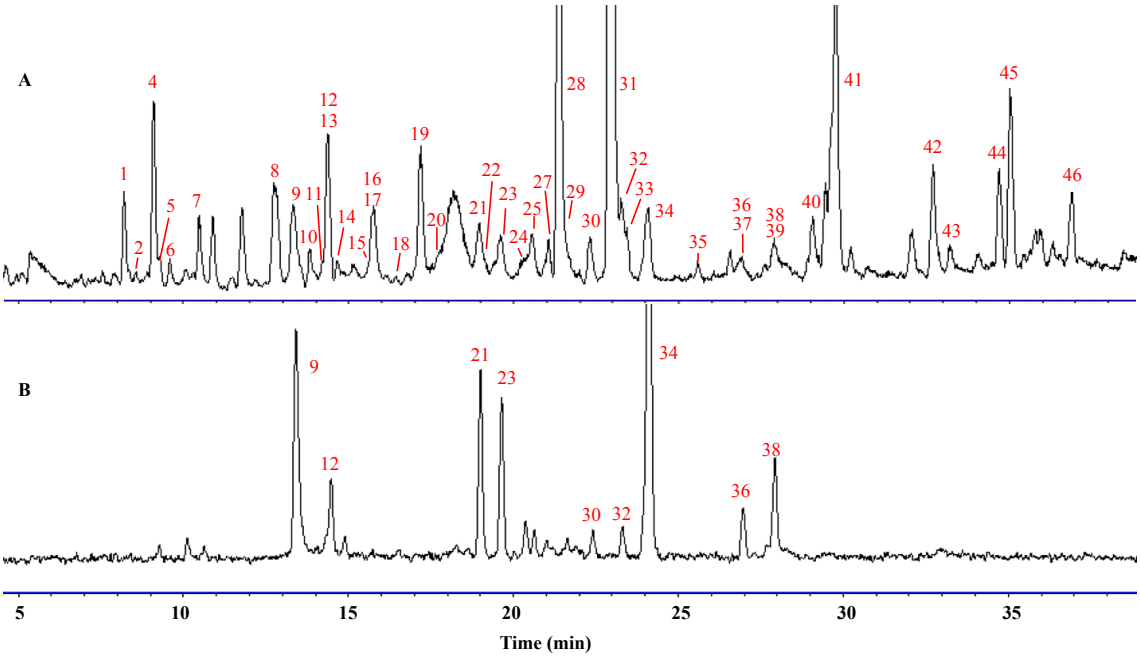


Figure VI-2. Representative chromatograms of alkaloids, flavonoids and saponins in *Solanum scabrum* SS 04.2. (**A**) total ion chromatogram (TIC) and (**B**) UV-vis chromatogram at 370 nm acquired by method (a). The identities, retention time and MS of each peak are listed in **Table VI-1**. Prefix "a-" in compound codes were not written for clarity.

Table VI-1. Identification of major compounds in berries of eight *Solanum scabrum* accessions

Code	RT (min)	Compound ID	Parent & fragmental ions (*m/z*) (CID-80 %)	Other major fragments (*m/z*) (CID-150 %)	λ max
a-1	8.1	DHS-hex-rha-hex	916, 754, 608, 446	428, 410, 287, 269, 251	-
a-2	8.3	DHS-hex-rha-rha	900, 754, 608, 446	428, 410, 287, 269, 251	-
a-3	8.8	DHS-hex-hex-rha	916, 770; 608; 446	428, 410, 287, 269, 251	-
a-4	9.1	DHS-hex-rha-rha	900, 754, 608, 446	428, 410, 287, 269, 251	-
a-5	9.3	DHMS-hex-rha-hex	930, 768, 622, 460	442, 424, 287, 269, 251	-
a-6	10.0	N ◊	377*, 355	181, 163	328, 260 (sh)
a-7	10.4	DHMS-hex-rha-rha	914, 768, 622, 460	442, 424, 287, 269, 251	-
a-8	12.8	HS-hex-hex-rha	900, 754, 592, 430	412.8, 271.5, 253.5	-
a-9	13.3	C ◊	377*, 355	181, 163	328, 260 (sh)
a-10	13.5	DHS-hex-rha-mal-rha	986, 840, 754, 608, 446	428, 410, 287, 269, 251	-
a-11	14.2	DHS-hex-rha-mal-rha	986, 840, 754, 608, 446	428, 410, 287, 269, 251	-
a-12	14.4	Q-hex-hex-rha	794*, 773, 627, 465, 303	285, 257, 153, 137	355, 268
a-13	14.4	HS-hex-rha-rha	884, 738, 592, 430	412, 394, 271, 253	-
a-14	14.6	HS-hex-rha-hex	900, 738, 592, 430	412, 394, 271, 253	-
a-15	15.6	DHMS-hex-rha-mal-rha	1000, 854, 768, 622, 460	442, 424, 287, 269, 251	-
a-16	15.7	HS-hex-rha-rha	884, 738, 592, 430	412, 394, 271, 253	-
a-17	15.8	HS-hex-rha-hex	900, 738, 592, 430	412, 394, 271, 253	-
a-18	16.1	DHMS-hex-rha-mal-rha	1000, 854, 768, 622, 460	442, 424, 287, 269, 251	-
a-19	17.1	HS-hex-rha-rha	884, 738, 592, 430	412, 394, 271, 253	-
a-20	17.7	HS-hex-rha-rha	884, 738, 592, 430	412, 394, 271, 253	-
a-21	18.9	Q-hex-rha-rha	779*, 757, 611, 465, 303	285, 257, 153, 137	355, 268
a-22	19.0	HMS-hex-hex-rha	914, 768, 606, 444	426, 271, 253	-
a-23	19.6	Q-hex-hex	649*, 465, 303	285, 257, 153, 137	355, 268
a-24	20.2	HS-hex-rha-mal-rha	970, 824, 738, 592, 430	412, 394, 271, 253	-
a-25	20.3	HMS-hex-rha-rha	898, 752, 606, 444	426, 271, 253	-
a-26	20.4	HS-hex-rha-mal-rha	970, 824, 738, 592, 430	412, 394, 271, 253	-
a-27	21.1	HS-hex-rha-mal-rha	970, 824, 738, 592, 430	412, 394, 271, 253	-
a-28	21.3	S-hex-rha-hex	884, 722, 576, 414	396, 271, 253	-
a-29	21.5	HS-hex-rha-rha	884, 738, 592, 430	412, 394, 271, 253	-
a-30	22.3	I-hex-hex	663*, 641, 479, 317	163, 153	355, 268
a-31	22.9	S-hex-rha-rha (solamargine) ◊	868, 722, 576, 414	396, 271, 253	-
a-32	23.2	I-hex-hex	663*, 641, 479, 317	163, 153	355, 268
a-33	23.3	HS-hex-rha-rha-mal	970, 884, 738, 592, 430	412, 394, 271, 253	-
a-34	24.0	Q-hex-rha	633*, 611, 465, 303	285, 257, 153, 137	355, 268
a-35	25.5	S-hex-rha	722, 576, 414	396, 271, 253	-
a-36	26.9	I-hex-rha	647*, 478, 317	163, 153	355, 268
a-37	27.0	HMS-hex-rha-mal-rha	984, 838, 752, 606, 444	426, 271, 253	-
a-38	27.9	I-hex-rha	647*, 478, 317	163, 153	355, 268
a-39	27.9	S-hex-rha-mal-hex	970, 808, 722, 576, 414	396, 271, 253	-

a-40	29.1	D-hex-hex-hex-hex-rha-rha	1355, 1209, 1063, 901, 739, 577, 415	397, 271, 253	-
a-41	29.8	S-hex-rha-mal-rha	954, 808, 722, 576, 414	396, 271, 253	-
a-42	32.7	D-hex-hex-hex-rha-rha	1193, 1047, 901, 739, 577, 415	397, 271, 253	-
a-43	33.2	D-hex-hex-hex-rha-rha	1193, 1047, 901, 739, 577, 415	397, 271, 253	-
a-44	34.7	T-hex-hex-hex-hex-hex	1249*, 1227, 1087*, 1065, 903, 741, 579, 417	399, 273, 255	-
a-45	35.1	T-hex-hex-hex-hex-hex	1249*, 1227, 1087*, 1065, 903, 741, 579, 417	399, 273, 255	-
a-46	36.9	D-hex-hex-rha-rha	1031, 885, 739, 577, 415	397, 271, 253	-
b-1	5.7	P-hex-rha-hex	787, 625, 479, 317	-	529, 280
b-2	6.4	M-hex-rha-hex	801, 639, 493, 331	-	530
b-3	12.0	P-(hex)-hex-hex-cou	949, 787, 479, 317	-	535, 310, 280
b-4	13.1	Dp-(hex)-rha-hex-cou	919, 757, 465, 303	-	535, 310, 280
b-5	13.6	P-(hex)-hex-rha-cou	933, 771, 479, 317	-	535, 310, 280
b-6	14.9	P-(hex)-hex-rha-cou	933, 771, 479, 317	-	535, 310, 280
b-7	17.0	M-(hex)-hex-rha-cou	947, 785, 493, 331	-	534, 310, 280
b-8	22.2	P-(hex)-rha-cou	771, 479, 317	-	544, 320

RT, retention time. For compound codes, "a-", compounds identified by method (a), and "b-" by method (b). For compound ID: N, neochlorogenic acid; C, chlorogenic acid; Q, quercetin; I, isorhamnetin; S, solasodine; HS, hydroxysolasodine or isomer; HMS, hydroxymethylsolasodine or isomer; DHS, dihydroxysolasodine or isomer; DHMS, dihydroxymethylsolasodine or isomer; D, diosgenin; T, tigogenin; P, petunidin; Dp, delphinidin; M, malvidin; hex, hexosyl; rha, rhamnosyl; xyl, xylosyl; mal, malonoyl, cou, coumaroyl. Side chain moieties conjugated on different aglycone locants were in parenthesis. $^{\diamond}$, confirmed by comparison with authentic standard. With respect to *m/z* and UV values: bold, protonated parent ions; bold with *, sodium adducted parent ions; *, sodium adducted fragmental ions; bold italicized, protonated aglycone ions; (sh), UV shoulder; "-", undetermined for CID-150 % fragments and lack of significant UV-vis absorption for λ_{max}.

3.1 Identification of phenolic acids.

Two major phenolic acids were detected as chlorogenic acid (**9**) and its isomer neochlorogenic acid (**6**), with the former being the predominant phenolic acid. Both compounds had UV maxima at 249 nm and 328 nm (**Figure A-14A**), which agreed with the literature [5]. With MS-CID 80%, both compounds featured strong sodium adducted parent ions of 377 *m/z* ($[M+Na]^+$) and moderate protonated ions of 355 *m/z* ($[M+H]^+$). With enhanced CID 150 %, the majority of parent ions were broken down into its components of quinic acid, which was almost invisible in the mass spectrum, and protonated caffeic

acid as weak peak at 181 *m/z*, which continued to lose a water molecule to form the base peak 163 *m/z*. The identification was confirmed by comparison with retention time and mass spectra of authentic standards.

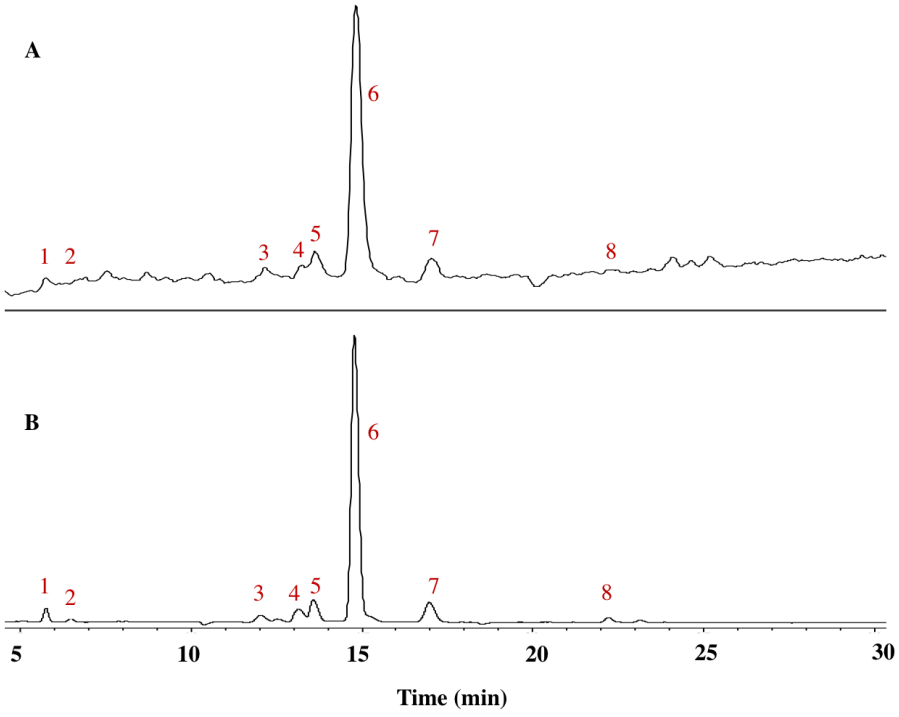


Figure VI-3. Anthocyanins in *Solanum scabrum* of USDA PI 643126 shown by (**A**) total ion chromatogram (TIC) and (**B**) UV-vis chromatogram at 520 nm as acquired by method (b). The identities, retention time and MS of each peak are listed in **Table VI-1**. Prefix "b-" in compound codes were not written for clarity.

3.2 Identification of flavonols.

Glycosides of two major flavonols of quercetin and isorhamnetin were detected. Compound **34,** for example, featured characteristic absorption peak centered around 355 nm and 268 nm (**Figure A-14B**), which suggested a flavonol structure [6]. In MS-CID 80 %, it had sodium adducted parent ions of 633 *m/z* ($[M+Na]^+$) and protonated parent ions of 611 *m/z* ($[M+H]^+$), which had a loss of rhamnosyl to produce fragments of 465 *m/z*

([M+H-rhamnosyl]$^+$) and then an additional loss of hexosyl to form aglycone ions of 303 *m/z*. This suggested a quercetin-rhamnosyl-hexoside. Using MS-CID 150%, most parent ions and glycosylated fragments were shattered into aglycone ions as the base peak. A small fraction of the aglycone ions was further broken into pieces of 257 *m/z* ([M+H-H_2O-CO]$^+$) and 229 *m/z* ([M+H-H_2O-2CO]$^+$) by rupture of B ring and C ring, along with fragments of 153 *m/z* ($^{1,3}A^+$) and 137 *m/z* ($^{0,2}B^+$) through characteristic retro Diels-Alder (RDA) cleavage of C ring, which agreed well with literature [7]. MS/MS of quercetin freed by acid hydrolysis revealed more structural evidence including fragment 165 *m/z* ($^{0,2}A^+$). The identity of the freed aglycone in hydrolyzed extract was further confirmed by comparison with quercetin standard.

Glycosides with the aglycone ion of 317 *m/z*, which is 14 *m/z* higher than protonated quercetin, was identified as glycosides of methylated quercetin (**Figure A-12**, **a-36**). The methyl substitution location, however, was difficult to identify based on MS with either CID 80% or 150 % because of suppressed aglycone fragmentation, which possibly was a result of interference from the methyl group [7]. Instead, MS/MS of the aglycone freed from the pool of corresponding glycosides after hydrolysis was shown to be more informative. The occurrence of fragment 165 *m/z* ($^{0,2}A^+$) suggested an A ring structure like that of quercetin, and fragment 151 *m/z* ($^{0,2}B^+$), which was 14 *m/z* higher than $^{0,2}B^+$ piece of quercetin, suggested a methyl substitution on the B ring. The aglycone identity was further confirmed to be isorhamnetin by comparison with standard. It is interesting to notice that the base peak was an even numbered 302 *m/z* formed by the loss of methyl radical. Despite the rarity of formation of radicals in ESI, the formation of radical ions could be rationalized by the extended conjugation in the molecules, and similar phenomena

could also be observed in other methylated flavonoids [8]. It is of interest to note that isomer rhamnetin was found in the leaves of *S. scabrum* with methyl substitution occurring on the A ring [9].

3.3 Identification of glycoalkaloids.

3.3.1 Identification of "414-series"

Identification of glycosylation pattern of glycoalkaloids was similar to that of flavonol glycosides previously discussed. Apart from glycosylation, multiple glycoalkaloids featured side chains with acylation by malonic acid as suggested by the neutral loss of 86 *m/z*. These acylated glycoalkaloids were eluted at the end of series of glycosides of the same aglycone. Conjugation of the malonic acid in the oligosaccharide chain in glycoalkaloids is rather uncommon, and rarity of report on such acylated compounds may result from the fact that the study of natural products has traditionally been dominated by column separation for NP purification, and the heat applied during rotavapping caused the malonyl piece to be cleaved off, [10] which was also observed in this study.

Regarding aglycone identification, the aglycone ion of 414 *m/z* could be readily categorized as a nitrogen-containing alkaloid based on its even number molecular mass [3], and was further identified as protonated solasodine based on the major fragmental ions $^{17\text{-}20,\,22\text{-}O}ABCD^{+}$ of 271 *m/z* and $^{17\text{-}20,\,16\text{-}O}ABCD^{+}$ of 253 *m/z* in MS^{1}-CID 150% (**Figure A-12**, **a-41**) and MS/MS experiment (**Figure VI-5**). These two characteristic fragments were formed by rupture of E ring of the aglycone via pathway shown in **Figure VI-6** [11, 12]. In addition, the fragmental ion of low abundance 157 *m/z* as the fragment derived from $^{17\text{-}}$

$^{20,16\text{-}O}ABCD^{+}$ of 253 *m/z* was also visible in MS/MS spectrum (pathway not shown) [12]. The identity of the aglycone freed after hydrolysis was further confirmed by comparison with authentic standard as solasodine, and glycosides of solasodine has been frequently reported as the major glycoalkaloids in many Solanaceae species including *S. scabrum* [13, 14]. Compound **a-31** was confirmed by comparison with standard to be solamargine, one of the most common glycoalkaloid in Solanaceae family.

Apart from glycosides of solasodine, series of glycoalkaloids with less common or potentially novel aglycone structures were also detected, and the structures were putatively identified by comparing the chromatographic behavior and mass shift pattern with that of solasodine counterparts as discussed below.

3.3.2 Identification of "430-series"

The aglycone ion of 430 *m/z* showed increased polarity as suggested by shorter retention time of corresponding glycosides and post-hydrolysis free aglycones (**Figure VI-4**) on reversed phase column, and the mass increment by 16 *m/z* indicated substitution of an additional hydroxyl group on solasodine. Compared with solasodine ion in MS-CID 150%, the aglycone ions of 430 *m/z* showed stronger dehydration products of 412 *m/z* as a result of this hydroxyl substitution. Fragmental ions of 394 *m/z* produced by loss of two hydroxyl groups were also visible but faint. Fragments 271 *m/z* and 253 *m/z* shared by solasodine suggested a spirostane structure with A-B-C-D ring similar to that of solasodine, and thus occurrence of the hydroxyl group on the E-F ring (**Figure VI-5-B**, **C** and **D**, and **Figure A-12**, **a-33**). The profiled 430 *m/z* aglycones freed by acidic hydrolysis, however, were resolved into three major peaks with similar MS/MS spectra but varied ratio of ion

abundance (**Figure VI-4**, and **Figure VI-5**-**B**, **C** and **D**), manifesting the existence of multiple isomers. This complexity was also revealed by existence of multiple glycosides of 430 *m/z* aglycone with the same oligosaccharide pattern, such as compounds **a-13**, **a-16**, **a-19** and **a-20**. Thus, the aglycone ion 430 *m/z* was tentatively identified as E-F ring substituted hydroxysolasodine (HS) including different constitutional isomers as well as stereoisomers.

3.3.3 Identification of "446-series"

The aglycone ion 446 *m/z* was 16 *m/z* higher than HS ion with higher polarity as suggested by elution order of corresponding glycosides and post-hydrolysis free aglycones (**Figure VI-4**). In addition, this aglycone featured two-step dehydration products of 428 *m/z* and 410 *m/z* of moderate abundance (**Figure A-12**, **a-1**). This aglycone was thus identified as dihydroxysolasodine (DHS). The fragmental ions 287 *m/z* and 269 *m/z* was 16 *m/z* higher than the corresponding ions of $^{17\text{-}20,\ 22\text{-}O}ABCD^{+}$ and $^{17\text{-}20,\ 16\text{-}O}ABCD^{+}$ characteristics of S and HS, respectively, suggesting one hydroxylation on the fused A-B-C-D ring while the other on the spiro E-F ring. Besides, the occurrence of dehydration fragments of 251 *m/z* derived from 269 *m/z* was additional evidence of hydroxylation on the A-B-C-D ring (**Figure VI-5** and **Figure A-12**, **a-1**).

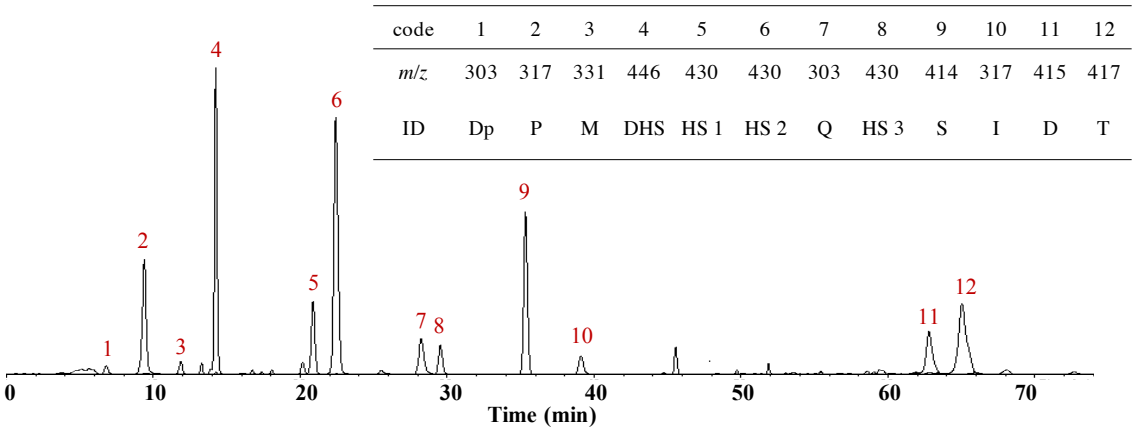


Figure VI-4. Extracted ion chromatogram (EIC) of aglycones freed by acid-assisted hydrolysis as acquired by method (c). Inset was a summary of aglycone identification. Compound abbreviations refer to **Table VI-1**.

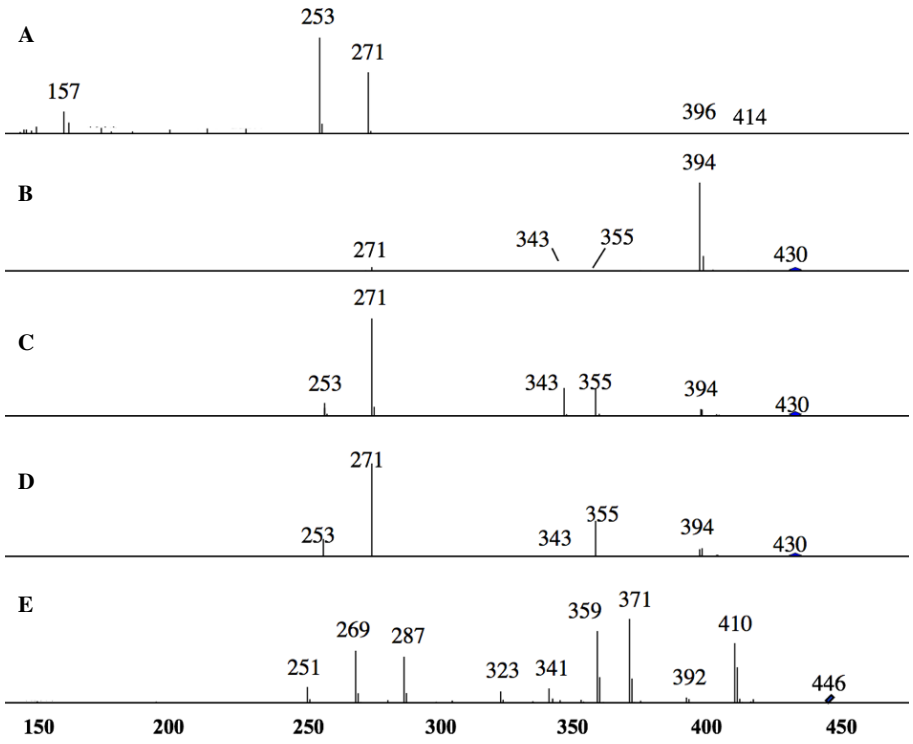


Figure VI-5. MS/MS spectra of post-hydrolysis free aglycones of glycoalkaloids acquired by method (c). The aglycones investigated were solasodine (**A**) (**9** in **Figure VI-4**); isomers of hydroxysolasodine or HS (**B**, **C** and **D**) (**5**, **6** and **8** in **Figure VI-4**); and dihydroxysolasodine or DHS (**E**) (**4** in **Figure VI-4**).

3.3.4 Identification of "444 & 460 series"

The aglycone ion of 444 *m/z* was 14 *m/z* higher than HS of 430 *m/z* with reduced polarity as suggested by elution sequence of corresponding glycosides, and along with characteristic 271 *m/z* and 253 *m/z* fragmental ions, it was tentatively identified as hydroxymethylsolasodine (HMS) with substitution of a methyl and a hydroxyl group on the E-F ring (**Figure A-12**, **a-37**). We suggest that it is also possible that the assumed methyl and hydroxyl group on the E-F ring could instead take a combined form as a methoxyl group. The aglycone ion of 460 *m/z* was 14 *m/z* higher than DHS ion of 446 *m/z* with decreased polarity, and featured similar two-step dehydration products and same $^{17\text{-}20,\ 22\text{-}O}ABCD^{+}$ and $^{17\text{-}20,\ 16\text{-}O}ABCD^{+}$ fragments as DHS (**Figure A-12**, **a-5**). This aglycone was thus identified as dihydroxymethylsolasodine (DHMS), with one hydroxyl group occurring on the A-B-C-D ring and the other hydroxyl and one methyl group on the E-F ring. As in the case of HMS, the assumed hydroxyl and methyl group on the E-F ring could instead take the combined form as a methoxy group.

Fragmental ions composed of nitrogenous F ring could be readily distinguished by the even mass number and provide supplemental information as to substitution pattern of the alkaloid aglycones. Compared with solasodine that had weak fragmental ions of 114 *m/z*, DHMS had fragmental ions of 144 *m/z* (**Figure A-12**, **a-5** and **a-41)**. This 30 *m/z* increment suggested additional substitution by hydroxyl and methyl groups (or a single methoxyl) on the E-F ring of DHMS, agreeing with the inferred structure aforementioned. Nitrogenous fragments, however, were found to be very weak and scarce in general and only provided limited information.

Although solasodine (25R-22α*N*) has been proposed as the skeletal basis of various derivatives, it is possible that the stereoisomer tomatidineol (25S-22β*N*), another *Solanum* alkaloid, exists as the alternative which was unfortunately not readily distinguishable from solasodine by mass spectrometry [15].

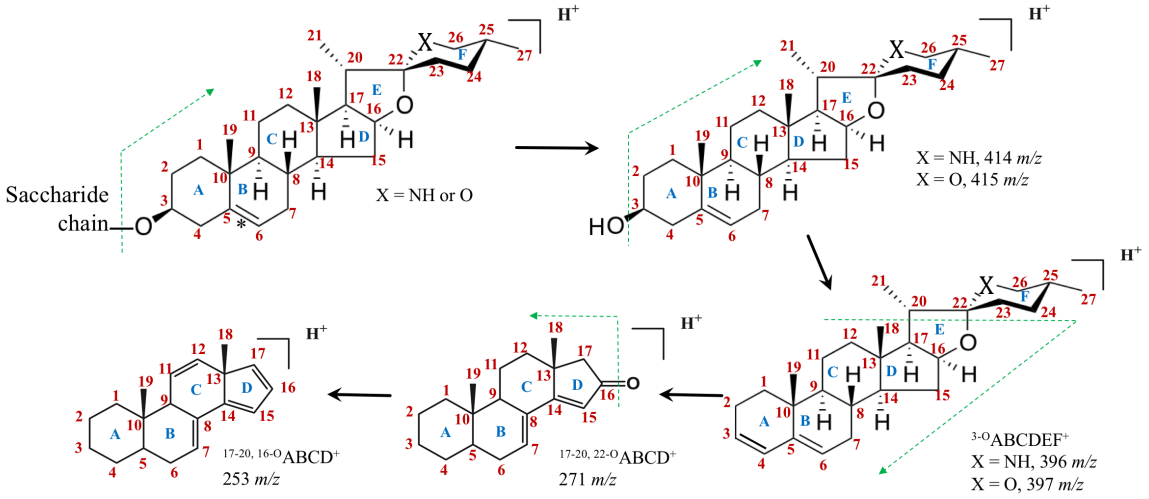


Figure VI-6. Characteristic fragmentation pathway of spirostan-derived alkaloid and saponin. *, saturation in the double bond position C5-C6, which leads to formation of fragments correspondingly 2 *m/z* higher. As to fragment nomenclature, for instance, the label $^{17\text{-}20,\ 22\text{-}O}$ABCD^{+} indicated fragmental ions containing the A-B-C-D rings, formed by the cleavage of bond between C17 and C20 (noted as "17-22"), and a second bond between C22 and the adjacent oxygen atom in ring E (noted as "22-O", "O" for oxygen atom).

3.3.5 Structural novelty

This is the first report that substituted solasodine derivatives are identified in *S. scabrum*. Such solasodine-substituted glycosides, which are far less common than the unmodified solasodine glycosides, have been occasionally reported in several other *Solanum* species. Hydroxylation has been the mostly reported substitution. Among E-F ring mono-hydroxylated solasodine, 23-hydroxysolasodine was first isolated as a bare aglycone from the roots of *S. canense* and *S. fraxinifolium* prepared with acid hydrolysis

[16] and the glycosylated form was later isolated from *S. uporo* [17]. Glycosides of 27-hydroxysolasodine (solaparnaine) were reported in unripe berries of *S. asperum* [18] and also found in chilled methanol-preserved two-year-old uncrushed berries of *S. nigrum* [19], the latter being a close species of *S. scabrum*. In addition, *N*-hydroxysolasodine was isolated from the roots of *Solanum robustum* [20]. For dihydroxylated solasodine, 12, 27-dihydroxysolasodine (C and F ring substitution) was isolated either as a bare aglycone or glycosides also from the berries of *S. nigrum* [19, 21]. Considering the rather scattered distribution of hydroxysolasodines across *Solanum* species, *S. scabrum* berries is indeed a rich source of such compounds; given the great compound diversity identified in this research as well as their moderate natural content, *S. scabrum* berries could be a new source for discovery of novel hydroxysolasodine glycosides.

Though methylated solasodine such as *N*-methylsolasodine was reported in *S. nigrum* [20], methylation-based modification on solasodine was indeed uncommon, and combination of methylation and hydroxylation (or in other aforementioned isomeric forms) have rarely, if any, been reported. Such rarity could be due to the low abundance in natural occurrence as noted in this study. Thus, the HMS and DHMS glycosides identified in this study could be potentially novel compounds. The uncommon and potentially novel glycoalkaloids detected in this study will be further subjected to purification and structural confirmation by nuclear magnetic resonance (NMR) in the future.

3.4 Identification of saponins

A group of glycosides of diosgenin and tigogenin were detected in the berries and these are almost universal in the *Solanum* species [1]. With distinctive long side chain,

tigogenin glycosides contained xylosyl moieties, a saccharide unit lacking in glycosides of other aglycones found in this study. Aglycone fragmentation pattern under MS-CID 150% was similar to that of corresponding alkaloids due to analogous structure. Diosgenin ion had fragmental ions $^{17\text{-}20,22\text{-}O}ABCD^{+}$ of 271 *m/z* and $^{17\text{-}20,16\text{-}O}ABCD^{+}$ of 253 *m/z* as shared by solasodine, HS and HMS. Tigogenin ion, saturated counterpart of diosgenin in the bond between C5-C6 featured fragments that were correspondingly 2 *m/z* higher, i.e., 273 *m/z* and 255 *m/z* (**Figure A-12**, **a-43**, **a-45**). The identity of aglycones, diosgenin and tigogenin were further confirmed by comparison with authentic standards after acidic hydrolysis.

3.5 Identification of anthocyanins.

For determination of anthocyanins, 0.4% trifluoroacetic acid (TFA) was used in replacement of 0.1 % formic acid as the mobile phase modifier to generate more acidic environment. By converting the pH-dependent structure of anthocyanidins to flavylium ions which had characteristic maxima absorption around 520 nm, this group of compounds could be readily identified and differentiated from isobaric flavonols by UV-vis signal [22]. We also found that TFA significantly improved peak shape versus formic acid (**Figure A-15**). In this study, a series of anthocyanins derived from aglycones of petunidin, delphinidin and malvidin were detected, which featured parallel pattern of glycosylation and acylation with coumaric acid, agreeing with the literatures [2, 23]. The compound **b-3**, for example, had an aglycone ion of 317 *m/z* corresponding to petunidin. The neutral loss of 162 *m/z* from 933 to 771 *m/z* and from 479 to 317 *m/z* indicated a hexose moiety. The 292 *m/z* difference between peaks 771 and 479 *m/z* indicated a simultaneous loss of either two rhamnosyls or a rhamnosyl and a coumaroyl. In comparison with non-acylated simple

petunidin anthocyanin **b-1** and petunidin standard, UV-vis profile of **b-3** showed additional maxima absorption around 310 nm suggesting acylation by a coumaroyl in the side chain (**Figure A-14-C**, **D** and **E**). Given that acylation generally occurs on the terminal saccharide, the first hexose that was cleaved off the parent ion (933 to 771 *m/z*) might be from a hydroxyl group (e.g. 5-OH) different from the one (e. g., 3-OH) conjugated with coumaroyl- rhamnosyl-hexose. Thus, for compound **b-3** we tentatively identified as petunidin-3-(*p*-coumaroyl-rutinoside)-5-O-glucoside, which has been reported as a major anthocyanin in the berries of *Solanum* and related species [23-25].

4 Conclusion

Three different and complementary HPLC-UV/Vis-MS or MS/MS methods were developed for identification of phytochemicals in *S. scabrum* berries: method (a) functioned as the major tool for phytochemical screening and identification, and method (b) was specifically tailed for analysis of anthocyanins, while method (c) complemented with acid-assisted hydrolysis allowed for further confirmation and elucidation of the aglycones in greater detail. Using all three methods, a total of 54 phytochemicals of polyphenols, glycoalkaloids and saponins in the berries from different genetic sources and maturity were identified. Particularly, a wide scope of toxic glycoalkaloids were reported for the first time, including less common or novel hydroxylated and/or methylated structures which were elucidated based on known scaffold-fragmentation pathway of MS data. This work, qualitative in nature, provides the foundation for future work that can provide quantitative determination of berry phytochemical profile and associated toxicity as well as nutrition value. In addition, the methods used in this work could apply to

facilitate quality control and contribute to screening or breeding for alkaloid-deficient genotypes that could serve as new food supply; or for alkaloid rich genotypes for extraction and further processing for industrial applications.

5 References

[1] Carle R. Investigations on the content of steroidal alkaloids and sapogenins within *Solanum* sect. *Solanum* (= sect. *Morella*)(Solanaceae). Plant Syst Evol 1981;138:61-71.

[2] Oszmianski J, Kolniak-Ostek J, Wojdylo A. Characterization of phenolic compounds and antioxidant activity of *Solanum scabrum* and *Solanum burbankii* berries. J Agric Food Chem 2014;62:1512-9.

[3] Johnson AR, Carlson EE. Collision-Induced Dissociation Mass Spectrometry: A Powerful Tool for Natural Product Structure Elucidation. Anal Chem 2015;87:10668-78.

[4] Ma YL, Li QM, VandenHeuvel H, Claeys M. Characterization of flavone and flavonol aglycones by collision-induced dissociation tandem mass spectrometry. Rapid Commun Mass Spectrom 1997;11:1357-1364. [In English]

[5] Chen HJ, Inbaraj BS, Chen BH. Determination of phenolic acids and flavonoids in Taraxacum formosanum Kitam by liquid chromatography-tandem mass spectrometry coupled with a post-column derivatization technique. Int J Mol Sci 2012;13:260-85.

[6] Zhen J, Villani TS, Guo Y, Qi Y, Chin K, Pan MH, Ho CT, Simon JE, Wu Q. Phytochemistry, antioxidant capacity, total phenolic content and anti-inflammatory activity of Hibiscus sabdariffa leaves. Food Chem 2016;190:673-680.

[7] Tsimogiannis D, Samiotaki M, Panayotou G, Oreopoulou V. Characterization of flavonoid subgroups and hydroxy substitution by HPLC-MS/MS. Molecules 2007;12:593-606.

[8] Wu Q, Wang M, Simon JE. Determination of isoflavones in red clover and related species by high-performance liquid chromatography combined with ultraviolet and mass spectrometric detection. J Chromatogr A 2003;1016:195-209.

[9] Yuan B, Byrnes D, Giurleo D, Villani T, Simon JE, Wu Q. Rapid screening of toxic glycoalkaloids and micronutrients in edible nightshades (*Solanum* spp.). J Food Drug Anal 2018;26:751-760.

[10] Coppin JP, Xu Y, Chen H, Pan M-H, Ho C-T, Juliani R, Simon JE, Wu Q. Determination of flavonoids by LC/MS and anti-inflammatory activity in *Moringa oleifera*. J Funct Foods 2013;5:1892-1899.

[11] Lelario F, Labella C, Napolitano G, Scrano L, Bufo SA. Fragmentation study of major spirosolane-type glycoalkaloids by collision-induced dissociation linear ion trap and

infrared multiphoton dissociation Fourier transform ion cyclotron resonance mass spectrometry. Rapid Commun Mass Spectrom 2016;30:2395-2406.

[12] Cahill MG, Caprioli G, Vittori S, James KJ. Elucidation of the mass fragmentation pathways of potato glycoalkaloids and aglycons using Orbitrap mass spectrometry. J Mass Spectrom 2010;45:1019-25.

[13] Ding X, Zhu F, Yang Y, Li M. Purification, antitumor activity *in vitro* of steroidal glycoalkaloids from black nightshade (*Solanum nigrum* L.). Food Chem 2013;141:1181-6.

[14] Adesina S, Gbile Z. Steroidal constituents of *Solanum scabrum* subsp. nigericum. Fitoterapia 1984;55:362-3.

[15] Milner SE, Brunton NP, Jones PW, O'Brien NM, Collins SG, Maguire AR. Bioactivities of glycoalkaloids and their aglycones from *Solanum* species. J Agric Food Chem 2011;59:3454-84.

[16] Ripperger H, Porzel A. (23S)-23-hydroxysolasodine from two *Solanum* species. Phytochemistry 1991;30:1299-1301.

[17] Ripperger H. Steroidal alkaloid glycosides from *Solanum* uporo. Phytochemistry 1997;44:731-4.

[18] Silva TMS, Camara CA, Freire KRL, Silva TGd, Agra MdF, Bhattacharyya J. Steroidal glycoalkaloids and molluscicidal activity of *Solanum asperum* Rich. fruits. J Braz Chem Soc 2008;19:1048-1052. [In English]

[19] Gu XY, Shen XF, Wang L, Wu ZW, Li F, Chen B, Zhang GL, Wang MK. Bioactive steroidal alkaloids from the fruits of *Solanum nigrum*. Phytochemistry 2018;147:125-131.

[20] Ripperger H, Porzel A. N-Hydroxysolasodine from Solanum robustum. Phytochemistry 1992;31:1837-1839. [In English]

[21] Yoshida K, Yahara S, Saijo R, Murakami K, Tomimatsu T, Nohara T. Changes caused by included enzymes in the constituents of *Solanum nigrum* berries. Chem Pharm Bull (Tokyo) 1987;35:1645-8.

[22] Xu Y, Simon JE, Welch C, Wightman JD, Ferruzzi MG, Ho L, Pasinetti GM, Wu Q. Survey of polyphenol constituents in grapes and grape-derived products. J Agric Food Chem 2011;59:10586-93.

[23] Price CW, Ronald E. Anthocyanin pigments of Royal Okanogan huckleberry juice. J Food Sci 1995;60:369-374.

[24] Andersen ØM, Opheim S, Aksnes DW, Frøystein NÅ. Structure of petanin, an acylated anthocyanin isolated from *Solanum tuberosum*, using homo-and hetero-nuclear two-dimensional nuclear magnetic resonance techniques. Phytochemical Analysis 1991;2:230-236.

[25] Francis F, Harborne J. Anthocyanins of the garden huckleberry, *Solanum guineese*. J Food Sci 1966;31:524-528.

CHAPTER VII. BERRY PHYTOCHEMICAL QUANTIFICATION

1 Introduction

This study focused on quantitation of nightshades berry phytochemicals identified in the prior CHAPTER V. The berries investigated in this work included multiple genetic sources and across different maturation stages cultivated under the same environment. Based on the achieved data, this study serves to provide guidance for nutritional and safety evaluation and comprehensive applications of an underutilized agricultural resource to increase food supply, improve nutrition and enhance income in sub-Saharan Africa.

2 Materials and Methods

2.1 Chemical reagents

Authentic reference standards of neochlorogenic acid, chlorogenic acid, quercetin and dioscin were purchased from Sigma-Aldrich (St. Louis, MO), isorhamnetin from Carl Roth (Karlsruhe, Germany), solamargine from MedChem Express (Monmouth Junction, NJ), delphinidin chloride, petunidin chloride and malvidin chloride from ChromaDex, Inc. (Irvine, CA). Methanol, trifluoroacetic acid (TFA), HPLC-MS grade formic acid, and HPLC grade water and acetonitrile were purchased from Fisher Scientific (Fair Lawn, NJ).

2.2 Plant Materials.

Seeds of a total of eight entries of *S. scabrum* were sourced, with six entries provided from the World Vegetable Center (WorldVeg), i.e., Ex Hai, BG 29, SS 52, BG 16

(Nduruma), SS 49 (Olevolosi) and SS 04.2, one entry from the U.S. Department of Agriculture (USDA) PI 643126, and one commercial entry from Baker Creek Heirloom Seeds, 'Garden Huckleberry', (Lot #333BC, Mansfield, MO).

Seeds were sown in 72-cell trays with growing mix (Fafard Grow Mix 2; Sun Gro Horticulture, Agawam, MA) and germinated at the New Jersey Agricultural Experiment Station (NJAES) Research Greenhouse of Rutgers University. After four weeks of growth, the seedlings were transplanted to raised beds covered with 1.25mm black plastic mulch with drip irrigation applied as needed in New Jersey Agricultural Experiment Station (NJAES) Horticultural Research Farm #3, New Brunswick, New Jersey. Berries of three different maturation stages were manually harvested in mid-September 2016, and categorized as immature berries, firm with green exterior; half-mature berries, purple exterior with green interior; and mature berries, soft with dark purple color in both the exterior and interior. Late harvested berries after occurrence of frost were collected in early November. The harvested berries were dried in an air-circulated 40 °C oven and ground using a shearing-action mill with 2 mm mesh filter screen on the out-flow. Ground samples were stored in dual-layer resealable bags in a cool place out of direct light.

2.3 Instrumentation and equipment.

For the phytochemical determination, the instrument used was Agilent 1100 series LC/MSD ion trap (Agilent Technologies, Waldbronn, Germany) equipped with a degasser, quaternary pump, column thermostat and diode array detector (DAD). Compound separation was achieved using column Polaris 180Å Amide-C18, 250 x 4.6 mm, 3 μm

(Santa Clara, CA). The software was Agilent ChemStation A.08.03 and LC/MSD Trap Control 5.1.

2.4 Sample preparation.

For phytochemical quantification by LC-MS, ~ 0.2 g samples were extracted by 25 mL 70 % methanol acidified with 0.1 % formic acid, vortexed and then sonicated for 20 min. Next, the extract was allowed to stand at room temperature overnight. The extract was then centrifuged at 16,000 ×g for 10 min, and the supernatant was ready for LC-MS injection.

2.5 Quantification of phenolic acids, flavonols, glycoalkaloids and saponins.

For determination of phenolic acids, flavonol glycosides, glycoalkaloids and saponins in *S. scabrum* berries, the mobile phase was water with 0.1 % formic acid (A) and acetonitrile with 0.1 % formic acid (B) with a flow rate of 1 $mL \cdot min^{-1}$. The gradient was 10 to 15 % B from 0 to 5 min, 15 to 30 % B from 5 to 35 min, 30 to 50 % B from 35 to 55 min, and 50 to 60 % B from 55 to 60 min, followed with equilibration using 10% B for 10 min before the next injection. The column thermostat was set at 25 °C. The injection volume was 10 μL. The diode array detector (DAD) was set at 210 nm, 254 nm and 370 nm, with 400 nm as reference wavelength. Approximately a third of HPLC eluent was diverted into MS. With respect to electrospray ionization (ESI), the nebulizer needle voltage was 3500 V. Nitrogen was used as nebulizing gas set at 40 psi and as drying gas at 350 °C with a flow rate of 10 $L \cdot min^{-1}$. Helium was used as the collision gas. The collision energy was set at 80 % for generation of the characteristic collection of parent ions,

glycosylated and/or acylated fragmental ions and the aglycone ions. The scanning mode was set at positive polarity and the scan range was 100~1500 *m/z*. Ion counts control (ICC) was set at targeted 40,000 and the maximum accumulation time was 300 ms.

Quantification of phytochemicals was achieved using representative reference standards of the corresponding category. For compounds containing characteristic chromophores, quantification was based by UV/Vis detection. Specifically, chlorogenic acid and neochlorogenic acid were quantified using the corresponding standards at 254 nm, and glycosides of quercetin and isorhamnetin were estimated at 370 nm by standards of the corresponding aglycones, i.e., quercetin and isorhamnetin, respectively, with the quantity corrected by the molecular weight ratio [1, 2]. For compounds lacking chromophores, i.e., the glycoalkaloids and saponins, quantification was based on MS detection using structurally representative standards of solamargine and dioscin, respectively, and the peak intensity was calculated as the sum of counts of the characteristic parent ions, glycosylated and/or acylated fragmental ions and the aglycone ions specified in prior research [3] and then corrected by the corresponding molecular mass ratio. The content was presented in SI unit of $g \cdot kg^{-1}$ dry weight (DW).

2.6 Quantification of anthocyanins.

For quantification of anthocyanins in *S. scabrum* berries, the experimental conditions remained the same as in the described method above except for the following specification. The modifier used for mobile phase A and B was 0.4 % trifluoroacetic acid (TFA) instead. The gradient was 13 to 17 % from 0 to 2 min, 17 to 20 % from 2 to 13 min, 20 to 30 %

from 13 to 25 min, and 30 to 50 % from 25 to 45 min. The column thermostat was set at 40 °C and the DAD was set at 520 nm.

Quantity estimation of anthocyanins was achieved using reference standards of the corresponding aglycones based on detection at 520 nm. Specifically, glycosides of delphinidin, petunidin and malvidin were estimated using external standards of delphinidin chloride, petunidin chloride and malvidin chloride, respectively, corrected by the molecular weight ratio [1, 2]. The content was presented in SI unit of $g \cdot kg^{-1}$ DW.

2.7 Data Analysis.

Data processing was achieved using Agilent DataAnalysis 2.2. Statistical analysis, visualization and annotation was achieved using Microsoft Excel 2016 and the R software suite. Specifically, comparison of the phytochemical profile within mature berries from eight varied genetic sources was conducted using R *ComplexHeatmap* and other packages [4, 5], where color assignment and Euclidean distance-based clustering analysis was applied upon logarithmically transformed data to accommodate the large variation in compound production. Comparison of phytochemical profile within berries of different stages of fruit development from two genetic sources were performed using two-way analysis of variance (ANOVA), and the correlation was analyzed using simple linear regression after respective standardization to z-score (difference between observed values and the mean then divided by standard deviation) for each compound category and each genetic source to facilitate visualization on the same scale.

3 Results and Discussion

3.1 Phytochemical profile of mature berries

Polyphenols, saponins and alkaloids in berries from different genetics and over differing stages of fruit maturation varied significantly in content (**Table VII-1** and **Table VII-2**). This could be seen in particular in mature berries, where the phytochemical profile was visualized and compared by heat map with accessions clustered based on profile similarity (**Figure VII-1**). There was a significant variation within mature berries in the accumulation of phenolic acids, 0.91 ~ 7.95 $g \cdot kg^{-1}$ dry weight (DW); flavonols 0.76~8.98 $g \cdot kg^{-1}$ DW; anthocyanins 1.78 ~ 46.53 $g \cdot kg^{-1}$ DW; glycoalkaloids 0.02 ~ 16.34 $g \cdot kg^{-1}$ DW; and saponins 0.82 ~ 6.06 $g \cdot kg^{-1}$ DW.

Polyphenols were generally the most abundant compounds found in highest concentrations, then saponins in mature berries, while the reservoir of glycoalkaloids featured the highest compound diversity and most dramatic and widest range of production.

Of the striking dissimilarities of chemical profiles observed across different sources of mature berries, the discrepancy in glycoalkaloid profile was the most noticeable. Berries from USDA PI 643126 and commercial Baker Creek Heirloom, for example, had limited numbers of glycoalkaloids with a total accumulated yield below 0.07 $g \cdot kg^{-1}$ DW. In contrast, berries of accession SS 04.2 was characterized with a wider number of glycoalkaloids, which accounted for up to 15.74 $g \cdot kg^{-1}$ DW. The two accessions BG 16 (Nduruma) and SS 49 (Olevolosi), newly developed by WorldVeg and released in Tanzania and Mali in 2011[6], also contained moderately high level of glycoalkaloid content close to 3 $g \cdot kg^{-1}$ DW. Apart from glycoalkaloids, the phytochemical profile difference could also be seen in the polyphenol distribution. The polyphenols in berries from WorldVeg SS 04.2., SS49

(Olevolosi) and BG 16 (Nduruma) were considered the most enriched and balanced between phenolics, flavonols and anthocyanins, while polyphenols in berries USDA PI 643126 and commercial Baker Creek Heirloom were mostly constituted by abundant anthocyanins, whereas those in WorldVeg Ex Hai and BG 29 were relatively lacking in both flavonols and anthocyanins. This dramatic profile difference reflected significant variation and complexity within the single species of *S. scabrum*, and further mirrored the complexity already realized in the section *Solanum* [7, 8]. Of importance to note is that the WorldVeg breeding and new crop program with this species focuses and selects for field performance only as a leafy green and not for any use with the berries.

Despite the remarkable variation of chemical portfolios across different sourced berries, noticeable similarity existed. All accessions contained nearly uniformly high contents of chlorogenic acid as well as neochlorogenic acid, along with almost homogenously distributed glycosides of diosgenin and tigogenin, which is in agreement with prior research [9]. Several phytochemicals with significant occurrence were also found to be shared in common across all accessions, such as compounds **a-34** and **a-38** (flavonols), **a-4**, **a-18** and **a-31** (glycoalkaloids) and **b-6** (anthocyanin), etc. The phytochemical fingerprint of *S. scabrum* berries could be used to facilitate source identification, authenticity examination, quality control and taxonomy.

Table VII-1. Content of phytochemicals in *Solanum scabrum* berries of different genetic resources and maturation stages.

Class	Compound identification	SS 52	BG 29	BG 16	SS 49	SS 04.2	Baker Creek	WorldVeg Ex Hai				USDA PI 643126			
		mature	mature	mature	mature	mature	mature	raw	medium	mature	late harvest	raw	medium	mature	late harvest
PA	a-6_N	16.9	17.9	29.3	26.2	77.6	15.8	81.3	16.5	12.0	53.8	82.5	29.6	13.7	9.26
	a-9_C	126	137	171	362	717	203	523	189	112	219	450	242	77.7	9.26
	subtotal	143	155	200	388	795	219	604	206	124	273	532	271	91.5	18.5
FG	a-12_Q-h2-r	41.5	-	44.0	51.2	91.2	-	52.1	42.5	-	-	-	-	-	59.6
	a-21_Q-h-r2	64.2	77.6	92.4	80.8	162	145	288	122	69.9	129	51.6	40.4	33.6	-
	a-23_Q-h2	44.6	-	47.0	63.6	118	-	-	-	-	-	-	-	-	28.4
	a-34_Q-h-r	99.4	145	173	160	394	339	337	145	120	146	48.2	30.1	34.0	28.7
	a-30_I-h2	10.3	-	10.5	15.0	20.3	-	-	-	-	-	-	-	-	10.3
	a-32_I-h2	12.4	-	8.62	18.7	21.4	-	-	-	-	-	-	-	-	-
	a-36_I-h-r	16.1	19.6	20.6	19.2	32.3	32.8	65.2	27.9	15.8	32.6	-	-	-	-
	a-38_I-h-r	24.2	31.9	29.0	38.4	57.4	68.7	49.2	29.0	23.1	22.5	-	8.64	8.77	-
	subtotal	313	274	425	447	898	585	792	367	228	330	99.8	79.1	76.3	127
AC	b-1_P-h-r-h	59.9	28.6	14.0	98.3	47.4	89.1	-	-	14.2	-	-	-	57.5	16.1
	b-3_P-(h)-h2-c	-	-	12.6	61.8	54.0	103	-	-	-	-	-	14.4	59.8	13.6
	b-5_P-(h)-h-r-c	197	81.5	43.6	85.5	93.4	348	-	14.4	19.3	-	-	62.2	251	47.4
	b-6_P-(h)-h-r-c	1130	589	244	1020	517	3550	-	39.4	98.7	36.7	17.6	213	2380	394
	b-8_P-(h)-r-c	64.7	47.9	13.2	60.0	53.1	49.3	-	-	21.4	-	-	24.5	46.7	13.0
	b-2_M-h-r-h	-	-	-	-	-	28.1	-	-	-	-	-	-	22.5	-
	b-7_M-(h)-h-r-c	75.5	48.7	30.6	77.2	45.5	301	-	-	12.1	-	-	36.7	194	28.9
	b-4_Dp-(h)-h-r-c	53.1	58.8	28.8	110	94.9	187	-	-	12.4	-	-	37.0	123	25.8
	subtotal	1580	854	387	1520	905	4650	-	53.8	178	36.7	17.6	388	3130	539
	Total polyphenol	2030	1280	1010	2350	2600	5460	1400	626	531	640	649	739	3300	685

	a-28_S-h-r-h	2.61	0.63	2.11	18.3	264	0.67	287	20.2	-	104	12.7	0.44	-	1.27
	a-31_S-h-r-r	9.13	2.16	12.6	44.0	861	1.20	1600	247	0.54	689	76.4	0.88	0.63	6.24
	a-35_S-h-r	-	-	1.01	0.29	0.44	0.35	-	-	-	-	0.23	0.26	-	-
	a-39_S-h-r-m-h	-	-	0.42	0.73	0.57	-	0.39	0.32	-	0.87	-	-	-	-
	a-41_S-h-r-m-r	1.75	0.56	1.38	13.7	242	-	356	33.7	-	114	9.90	-	-	-
	a-8_HS-h2-r	2.89	4.66	9.76	14.1	19.9	0.51	8.30	15.8	1.32	24.1	1.11	0.49	-	-
	a-13_HS-h-r2	5.10	6.01	17.7	14.8	33.6	0.25	59.2	121	1.80	93.3	3.02	-	-	-
	a-14_HS-h-r-h	-	-	0.70	1.79	4.70	-	4.12	1.74	-	4.58	0.32	-	-	-
	a-16_HS-h-r2	0.40	0.33	0.94	2.10	12.2	-	20.8	5.97	-	3.37	1.05	-	-	-
	a-17_HS-h-r-h	0.39	0.33	0.39	2.65	6.25	-	-	-	-	-	0.22	-	-	-
	a-19_HS-h-r2	1.74	1.93	3.45	9.00	22.5	-	16.6	11.2	0.31	4.89	1.10	-	-	0.41
	a-20_HS-h-r2	-	-	-	-	1.73	-	2.30	1.23	-	-	-	-	-	-
	a-24_HS-h-r-m-r	-	-	-	-	1.22	-	2.32	3.87	-	1.76	0.29	-	-	-
	a-26_HS-h-r-m-r	-	-	-	-	-	-	3.88	5.67	-	1.97	0.38	-	-	-
	a-27_HS-h-r-m-r	0.81	0.51	1.16	3.09	6.19	-	10.9	19.2	0.40	14.9	1.08	-	-	-
GA	a-29_HS-h-r2	-	-	-	2.18	2.43	-	1.64	1.27	-	3.77	-	-	-	-
	a-33_HS-h-r2-m	1.00	0.80	1.15	8.03	24.6	-	18.5	15.0	-	-	1.06	0.28	-	-
	a-1_DHS-h-r-h	6.71	14.7	5.76	28.3	12.8	0.74	1.95	6.17	8.32	18.8	0.48	0.55	-	0.65
	a-2_DHS-h-r2	1.79	2.62	5.90	2.77	0.94	0.13	1.47	3.60	4.47	12.7	-	-	-	-
	a-3_DHS-h2-r	0.31	-	-	0.32	-	-	-	-	-	-	-	-	-	-
	a-4_DHS-h-r2	22.1	37.4	120	66.4	31.7	1.83	20.8	57.6	12.4	255	2.41	0.78	0.29	1.07
	a-10_DHS-h-r-m-r	0.62	0.83	1.82	2.10	1.55	-	0.82	1.67	1.04	5.68	-	-	-	-
	a-11_DHS-h-r-m-r	2.00	3.89	6.68	7.75	3.89	-	1.57	4.31	2.90	27.7	-	-	-	-
	a-22_HMS-h2-r	0.36	0.53	0.63	1.12	1.41	-	-	-	-	-	-	-	-	-
	a-25_HMS-h-r2	2.66	5.23	5.29	3.98	4.87	-	-	-	1.55	2.78	-	-	-	-
	a-37_HMS-h-r-m-r	1.10	1.38	1.03	1.44	0.92	-	-	-	0.65	1.67	-	-	-	-
	a-5_DHMS-h-r-h	1.80	4.56	9.79	4.45	1.68	-	-	1.68	5.05	10.9	-	0.25	-	-
	a-7_DHMS-h-r2	17.1	14.7	82.3	18.1	10.4	0.65	0.68	-	65.3	214	-	0.51	0.17	-
	a-15_DHMS-h-r-m-r	0.72	0.75	1.02	0.65	0.43	-	-	-	1.47	4.44	-	-	-	-
	a-18_DHMS-h-r-m-r	1.54	1.76	2.18	1.52	0.77	1.04	-	0.36	3.51	19.7	-	0.48	0.67	-
	subtotal	84.6	106	295	274	1570	7.37	2420	578	111	1630	112	4.92	1.76	9.64

	a-40_D-h4-r2	23.5	12.9	26.6	39.2	32.5	257	5.30	17.2	22.3	19.6	213	399	67.6	183
	a-42_D-h3-r2	7.72	5.25	7.08	17.9	53.3	26.6	7.27	8.11	2.87	20.5	84.8	3.41	-	17.7
	a-43_D-h3-r2	2.48	3.05	10.3	4.96	11.3	13.9	9.24	9.86	4.49	9.58	88.6	53.8	11.1	101
SA	a-46_D-h2-r2	2.87	-	3.07	6.83	13.3	16.4	80.7	10.8	3.85	8.63	111	5.44	3.60	14.6
	a-44_T-h5	18.0	19.4	84.9	13.4	45.0	58.4	66.5	50.8	19.2	311	29.3	29.8	19.0	228
	a-45_T-h5	27.6	50.5	294	108	174	233	170	115	65.5	262	47.0	157	46.3	429
	subtotal	82.2	91.1	425	191	330	606	339	211	118	631	573	648	148	973

Unit in mg·100 g-1 dry mass. For compounds category, PA, phenolic acids; FG, flavonol glycosides; AC, anthocyanins; PP, polyphenol; GA, glycoalkaloids; SA, saponins. For compounds identification, N, neochlorogenic acid; C, chlorogenic acid; Q, quercetin; I, isorhamnetin; S, solasodine; HS, hydroxysolasodine isomers; HMS, hydroxymethylsolasodine or methoxysolasodine; DHS, dihydroxysolasodine; DHMS, dihydroxymethylsolasodine or hydroxymethoxysolasodine; D, diosgenin; T, tigogenin; P, petunidin; Dp, delphinidin; M, malvidin; h, hexosyl; r, rhamnosyl; c, coumaroyl; m, malonoyl; cou, coumaroyl. Numbers immediately following "h" and "r" indicates the number of saccharide units connected. Side chain moieties conjugated on different aglycone locants were in parenthesis. For content notation, Tr, trace level, below 0.01 g·kg-1; "-", not detected. For germplasms, Entries Ex Hai, BG 29, SS 52, BG 16 (Nduruma), SS 49 (Olevolosi) and SS 04.2 were sourced from the World Vegetable Center (WorldVeg), PI 643126 from the U.S. Department of Agriculture (USDA), and one commercial line from Baker Creek Heirloom Seeds "Garden Huckleberry" (Lot #333BC, Mansfield, MO)

Table VII-2. Content of phytochemicals in *Solanum nigrum* berries of different genetic resources and maturation stages. The content is in units of mg / 100 g dry weight (DW)

Class	Compound identification	USDA PI 381289	Simlaw (Kenya)	USDA_PI 306400		
		mature	mature	raw	medium	mature
PA	a-6_N	14.05	11.74	17.77	30.10	20.13
	a-9_C	165.35	64.52	103.91	339.19	260.03
	subtotal	179.40	76.27	121.68	369.29	280.16
FG	a-12_Q-h2-r	0.00	0.00	0.00	0.00	0.00
	a-21_Q-h-r2	173.83	0.00	0.00	0.00	0.00
	a-23_Q-h2	0.00	47.66	64.98	131.04	89.81
	a-34_Q-h-r	0.00	0.00	0.00	0.00	0.00
	a-30_I-h2	9.65	8.95	16.81	39.18	34.39
	a-32_I-h2	0.00	0.00	14.22	26.11	23.21
	a-36_I-h-r	36.66	0.00	0.00	0.00	0.00
	a-38_I-h-r	91.16	0.00	0.00	0.00	0.00
	subtotal	311.30	56.62	96.02	196.33	147.40
AC	b-1_P-h-r-h	105.58	0.00	0.00	0.00	0.00
	b-3_P-(h)-h2-c	104.40	0.00	0.00	0.00	0.00
	b-5_P-(h)-h-r-c	288.92	0.00	0.00	0.00	0.00
	b-6_P-(h)-h-r-c	3301.82	0.00	0.00	0.00	14.93
	b-8_P-(h)-r-c	51.35	0.00	0.00	0.00	0.00
	b-2_M-h-r-h	25.88	0.00	0.00	0.00	0.00
	b-7_M-(h)-h-r-c	312.35	0.00	0.00	0.00	0.00
	b-4_Dp-(h)-h-r-c	152.24	0.00	0.00	0.00	0.00
	subtotal	4342.54	0.00	0.00	0.00	14.93
	Total polyphenol	4833.24	132.88	217.70	565.62	442.50
GA	a-28_S-h-r-h	0.00	0.37	21.58	3.22	0.35
	a-31_S-h-r-r	0.37	1.24	8.37	4.09	0.61
	a-35_S-h-r	0.00	0.17	0.33	0.00	0.00
	a-39_S-h-r-m-h	0.00	0.00	1.58	0.73	0.00
	a-41_S-h-r-m-r	0.00	0.40	11.62	2.39	0.00
	a-8_HS-h2-r	0.00	1.70	8.38	8.75	0.79
	a-13_HS-h-r2	0.00	1.31	9.45	6.47	0.63
	a-14_HS-h-r-h	0.00	0.00	0.00	0.00	0.00
	a-16_HS-h-r2	0.00	0.00	0.00	0.00	0.00
	a-17_HS-h-r-h	0.00	0.00	5.11	1.41	0.00
	a-19_HS-h-r2	0.00	0.32	9.67	2.56	0.42
	a-20_HS-h-r2	0.00	0.00	0.00	0.00	0.00
	a-26_HS-h-r-m-r	0.00	0.00	0.63	0.00	0.00
	a-27_HS-h-r-m-r	0.00	0.54	4.01	5.81	0.44
	a-24_HS-h-r-m-r	0.00	0.00	0.43	0.34	0.00
	a-29_HS-h-r2	0.00	0.00	2.08	0.00	0.00
	a-33_HS-h-r2-m	0.00	0.00	9.74	2.17	0.32

	a-1_DHS-h-r-h	0.28	23.93	17.60	19.61	13.19
	a-2_DHS-h-r2	0.00	0.00	0.00	0.00	0.00
	a-3_DHS-h2-r	0.00	0.68	0.62	1.91	1.64
	a-4_DHS-h-r2	0.78	17.22	18.91	0.21	2.64
	a-10_DHS-h-r-m-r	0.00	1.10	1.18	2.08	1.52
	a-11_DHS-h-r-m-r	0.00	4.60	5.66	20.16	10.70
	a-22_HMS-h2-r	0.00	0.00	1.82	4.90	4.39
	a-25_HMS-h-r2	0.00	0.76	4.06	9.99	11.61
	a-37_HMS-h-r-m-r	0.00	0.39	3.44	12.19	13.49
	a-5_DHMS-h-r-h	0.00	2.66	3.43	2.07	71.25
	a-7_DHMS-h-r2	0.35	8.15	6.27	8.33	72.94
	a-15_DHMS-h-r-m-r	0.00	0.42	0.57	2.35	4.32
	a-18_DHMS-h-r-m-r	1.01	1.61	3.39	25.19	53.91
	subtotal	2.79	67.57	159.93	146.91	265.16
SA	a-40_D-h4-r2	246.46	26.68	14.11	28.20	25.57
	a-42_D-h3-r2	29.18	0.00	15.12	15.83	4.66
	a-43_D-h3-r2	13.60	4.14	2.31	8.08	11.50
	a-46_D-h2-r2	16.43	3.18	3.64	14.47	8.75
	a-44_T-h5	48.75	0.00	16.73	8.25	1.83
	a-45_T-h5	225.65	14.82	17.18	148.95	112.59
	subtotal	580.08	48.81	69.08	223.78	164.89

Compound abbreviations refer to **Table VII-1**.

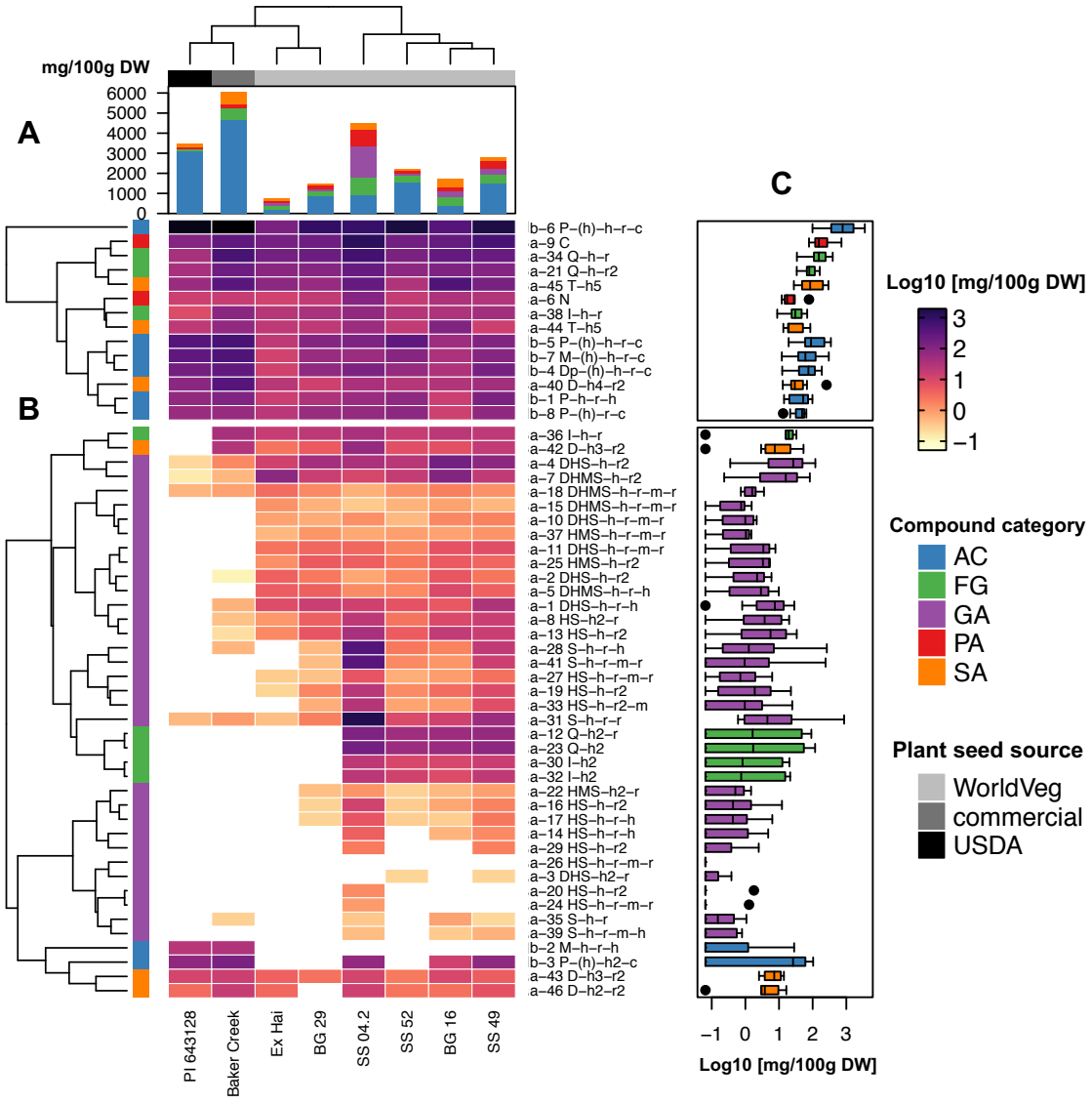


Figure VII-1. Phytochemical content in mature berries of *Solanum scabrum*. The stacked bar plot (**A**) shows categorized phytochemical subtotal in berries of different genetic sources. Clustered heatmap (**B**) presents compound distribution pattern and level of similarity. The boxplot (**C**) indicates the distribution of each compound across different genetic sources. (**A**), row side bar of (**B**), and (**C**) applies the same compound category color assignment. Compound abbreviations refer to **Table VII-1**. Note that (**B**) and (**C**) are presented in logarithmic scale.

3.2 Evolving phytochemical profile from immature to mature to post-frost.

The phytochemical portfolio in berries from WorldVeg Ex Hai and USDA PI 643126 featured dynamic changes over different fruit development stages, i.e., from immature to mature and finally to post-frost stage as presented in **Table VII-1** and **Figure VII-2**. The variation in accumulated content relative to magnitude of change over time in polyphenols, glycoalkaloids and saponins could be considered in part due to genetics and growth stage, as shown in **Figure VII-3**. The content and change in phenolic acids, for example, is a reflection more of a growth trait (explaining ~ 88% of total variation) more than built-in genetic differences, while dynamics in flavonol glycosides is more dominated by genetic singularity (accounting for ~ 54% total variation).

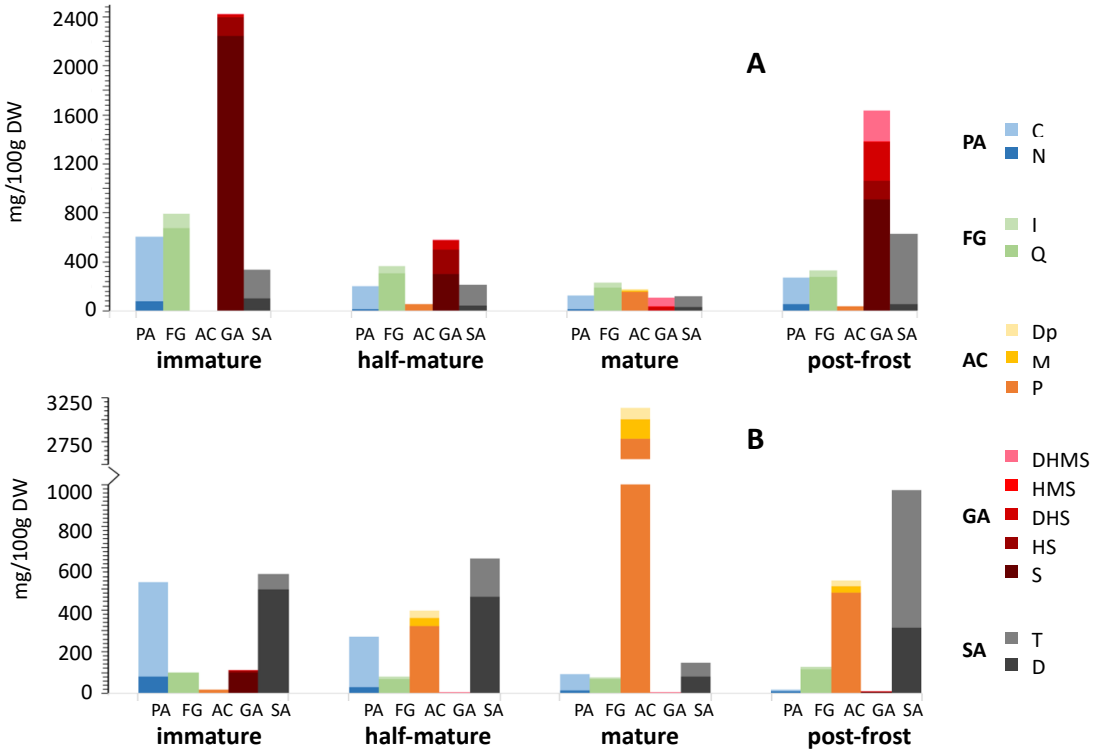


Figure VII-2. Phytochemical profiles of fruits of different development stages. (**A**) berries were sourced from WorldVeg Ex Hai, and (**B**) from USDA PI 643126. The abbreviations of analytes refer to **Table VII-1**.

Despite differences in absolute content, both accessions exhibited tendency for positively correlated trajectory in chemical portfolio change at different stages of fruit development as presented in **Figure VII-4**. Generally, during the course of fruit development and by the time of harvest in September, anthocyanins in both sourced berries showed steady accumulation, accompanied with color transition from green to dark purple both exterior and interior. The content of other compounds including phenolic acids, flavonols, glycoalkaloids and saponins all decreased by differing extent. Berries in the last post-mature harvest collected after frost in early November, however, featured reversed evolution in the phytochemical landscape: contents of anthocyanins decreased, while majority of other compounds "bounced back" to higher new level. Such increase in these second metabolites in the delayed harvest may be a result of protective response to cold stress, though the decrease in anthocyanins in this study somehow remained an exception [10, 11].

Among all the compounds, the change of glycoalkaloids was the most significant and dramatic. Glycosides of solasodine, predominant in raw green berries, diminished rapidly and drove the total collection of glycoalkaloids to diminution by ten to a hundred multifold by maturity in September. Though this phenomenon has been well noted [9, 12], it is indeed only a fraction of the complex evolution of the overall glycoalkaloid profile during entire growth stage as revealed in this study for the first time. In Ex Hai, as a typical example, glycosides of hydroxysolasodine (HS) and dihydroxysolasodine (DHS) featured "up-and-down" bell-curved change in content, while dihydroxymethylsolasodine (DHMS) and hydroxymethylsolasodine (HMS) including their potential methoxylated isomers, though HMS being low in concentration, exhibited gradual increment over the course of fruit

development. A bioconversion may be possible from solasodine to its hydroxylated counterparts and eventually to the hydroxymethylated or methoxylated forms by the time of maturity. Delayed harvest in November appeared to contribute to more complexity during which time glycoalkaloids of all types of aglycones increased by nearly fifteen times higher than the level at maturity in September. Cold stress is a likely trigger for inducement of glycoalkaloids as observed in the late harvest. These results are in agreement with the findings in other *Solanum* plants; *Solanum tuberosum* or potatoes were observed to contain marked increase in solanine after frost [13], a common potato glycoalkaloid of solanidane derivative (vs. spirosolane derivatives in *S. scabrum* berries).

Saponins, another major targeted group of bioactive natural products showed similar declining tendency during ripening but rebounded back to the highest level in delayed harvest. This tendency was accompanied with continuously increasing ratio of tigogenin glycosides in total saponin portfolio.

Figure VII-3. Source of contribution to the variation in phytochemical subtotal in fruits of different stages of development sourced from WorldVeg Ex Hai and USDA PI 643126. Sources of variations are attributed to main effects of genetic uniqueness, characteristics

of each specific maturation stage, and other effects (i.e., the residual term in two-way ANOVA analysis). Due to single-replication nature, the interaction of the two factors are merged into residual term.

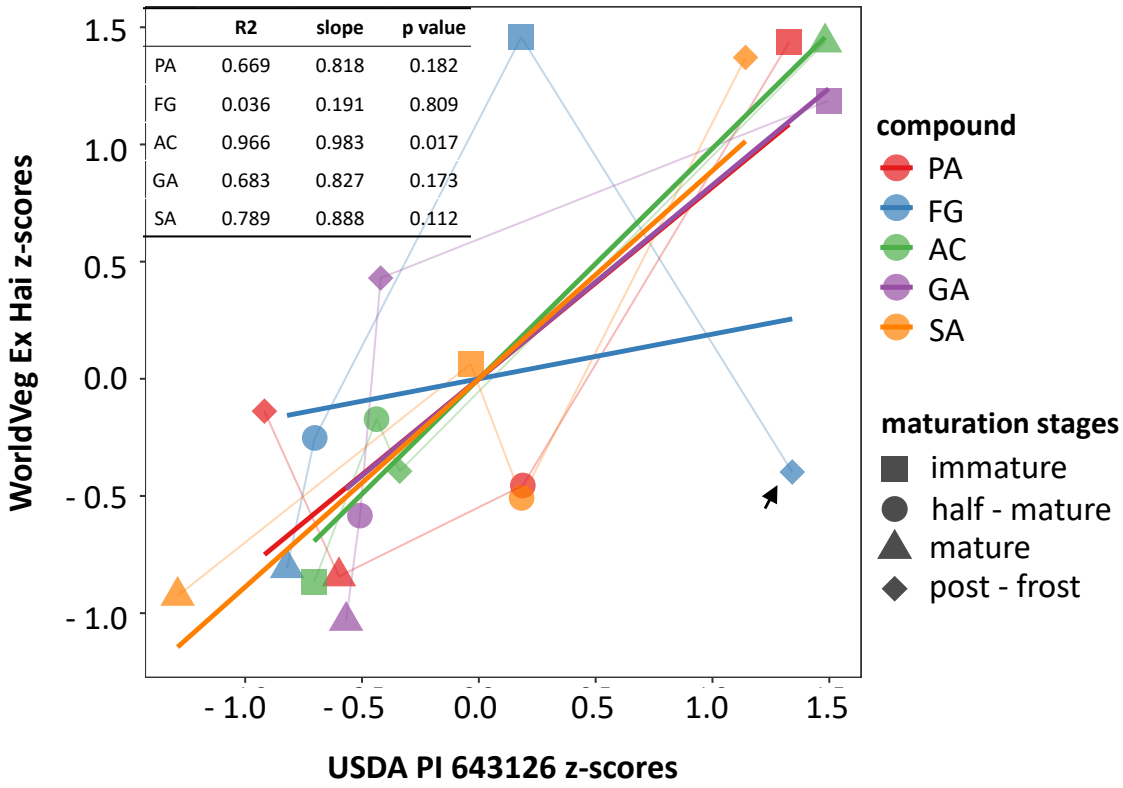


Figure VII-4. Correlation of phytochemical subtotal in fruits from WorldVeg Ex Hai and USDA PI 643126 during different stages of berry development. The accumulation content for each phytochemical category and generic source are standardized into z-scores, respectively. Notice the post-frost outlier for flavonol glycosides (FG) that leveraged flat the regression line. Points deriving from the same compound category are connected with faint lines to facilitate visual checking. Compounds abbreviations refer to **Table VII-1**.

3.3 Nutritional value from berry polyphenols

Dietary polyphenols are an important large group of natural products that provide multiple therapeutic effects against cancer, neuronal diseases, cardiovascular illnesses, diabetes, inflammation, and many such others diseases, and are playing an increasingly more important role in public health [14, 15]. Mature berries of *S. scabrum* were found to

be rich sources of such dietary antioxidants with anthocyanins being the most important constituents. As for total anthocyanins, in particular, except a few entries such as WorldVeg Ex Hai which contained relatively low content down to 1.78 $g \cdot kg^{-1}$ DW and was in the same magnitude reported by Oszmiański et al. (~2.06 $g \cdot kg^{-1}$ DW) [1], most entries contained times higher content close to or above 10 $g \cdot kg^{-1}$ DW. The highest content was seen in the entry marketed by Baker Creek Heirloom with a staggering level up to 46.53 $g \cdot kg^{-1}$ DW and USDA PI 643126 with 31. 28 $g \cdot kg^{-1}$ DW. Such remarkable levels, possibly due to targeted breeding or commercial selection, was in agreement with the magnitude reported by Cornelia et al. at 35 $g \cdot kg^{-1}$ DW [16] (assuming 90% moisture), comparable to or even higher than the level in the anthocyanin-dense blueberries at 6.15~32.06 $g \cdot kg^{-1}$ DW [17, 18] (assuming 90% moisture).

The abundant anthocyanins in mature berries could be of increased value nutritionally when consumed fresh (presuming low to no glycoalkaloids present) or extracted as food colorants and antioxidants. The largely acylated anthocyanins, which constituted 92 ~ 97 % of total anthocyanins as also reported in literature [1, 16, 19], exhibits elevated stability with heat, pH and light, and thus wider range of industrial application than simple anthocyanins in many other plant sources [20, 21].

3.4 Glycoalkaloid and saponin associated toxicity

Presence of toxic glycoalkaloids have been a major concern for consumption of many *Solanum* species including *S. scabrum* berries in sub-Saharan area [7, 9], though the berries are consumed in some other cultures. The differing rang in accumulation of glycoalkaloids across difference sources of berries may explain the cultural controversy on the edibility

of *S. scabrum* berries. With reference to the Food and Agriculture Organization / World Health Organization of the United Nations (FAO/WHO) standard for potato glycoalkaloids where the safety content threshold is 0.5 $g \cdot kg^{-1}$ DW [22] (assuming 80 % moisture content in fresh potato), and the reported safe consumption of household vegetable *Solanum melongena* or eggplant where solasodine glycosides occurred in a range of 0.63 ~ 2.05 $g \cdot kg^{-1}$ DW [23, 24] (assuming 90 % moisture content in fresh eggplant), the content of glycoalkaloids in berries WorldVeg SS 04.2 (15.74 $g \cdot kg^{-1}$ DW)., for instance, was 8~31 times higher than the reference thresholds and would be a health and safety concern. In fact, glycoalkaloid levels of this magnitude would already impart remarkable bitterness, compared with the potato glycoalkaloids perceivable threshold at 1.4 $g \cdot kg^{-1}$ FW (0.7 $g \cdot kg^{-1}$ DW, assuming 80% moisture content. Notice the closeness of perceivable threshold to abovementioned safety threshold set by FAO/WHO) and burning sensation threshold at 0.22 $g \cdot kg^{-1}$ FW (1.10 $g \cdot kg^{-1}$ DW) [25]. In contrast, the glycoalkaloids in berries from USDA PI 643126 with a significantly low content of less than 0.02 $g \cdot kg^{-1}$ DW, is not likely to cause safety concerns nor perceivable bitterness.

Removal of toxic glycoalkaloids, if necessary, may be achieved using solvent partition by their different solubility in acidic and alkaline environment or the practice of column chromatography to avoid the use of organic solvents [26, 27]. In village practice, processing with edible clay, depending on usage purpose and if possible, might be helpful for removal of glycoalkaloids by taking advantage of clay adsorption functionality, a practice traceable to geophagy arguably beneficial for enhancement of tolerance of glycoalkaloid-containing toxic potatoes [28].

Saponins are another major group of compounds whose toxicity have also been extensively investigated, and derivatives of triterpenes as found in the *S. scabrum* berries are generally considered toxicologically tolerable [29, 30]. Specifically, saponins of diosgenin and its structural analogues have been assessed to be safe for consumption at 0.267 $g \cdot kg^{-1}$ human body weight (BW) $\cdot day^{-1}$ in a most recent study by Xinxin et al. [31]; and dioscin, a most common saponin of diosgenin, was evaluated safe at 0.15 $g \cdot kg^{-1}$ animal $BW \cdot day^{-1}$ [32] or at an equivalent dose of 0.024 $g \cdot kg^{-1}$ human $BW \cdot day^{-1}$ [33], which corresponds to safe daily consumption for a person of 60 kg of around 240 g dehydrated mature berries or theoretically 2.4 kg fresh fruit of the most saponin-dense type (Baker Creek Heirloom). Based upon these reports and guidelines we can conclude the low concentrations of saponins in all the mature berries investigated would not be a safety concern for consumption on a reasonable daily amount.

Berries from a number of *S. scabrum* accessions were found to be of value for consumption, such as the commercial entry from Baker Creek Heirloom and USDA 643126, both of which had lowest and safe level of glycoalkaloids and saponins as well as the most enriched content of polyphenols. Particularly, in view of the highlighted consumption safety and nutrient value in the leaves from USDA 643126 [34], this accession could be a quality candidate to be introduced to and promoted in sub-Saharan Africa as an African indigenous vegetable, whose both fresh fruits and leaves are safe and nutrient-dense to consume.

4 Conclusion

The quantity of a total of 54 phytochemicals including polyphenols, glycoalkaloids and saponins in *S. scabrum* berries of different genetic origins and maturation stages were determined for the first time using a combination of two different HPLC-MS methods with representative reference standards. During the course of fruit maturation there were dynamic changes in phytochemical composition. Phytochemicals in mature berries from different genetic sources, apart from traits shared in common, had remarkable differences, and the genotype-dependent variance in toxic glycoalkaloids content may in part explain the berry consumption controversy in different cultural practices or geographical regions. In addition, mature berries from certain selected genetic sources had "safe" levels of glycoalkaloids and saponins as well as enriched polyphenols, and could be introduced to and promoted in sub-Saharan Africa as "new" indigenous crops, in which both leaves and fruits can be safely consumed and marketed at the village level. This could contribute to alleviate hunger and increase income in rural communities of sub-Saharan area.

5 References

[1] Oszmianski J, Kolniak-Ostek J, Wojdylo A. Characterization of phenolic compounds and antioxidant activity of *Solanum scabrum* and *Solanum burbankii* berries. J Agric Food Chem 2014;62:1512-9.

[2] Gasperotti M, Masuero D, Vrhovsek U, Guella G, Mattivi F. Profiling and accurate quantification of *Rubus* ellagitannins and ellagic acid conjugates using direct UPLC-Q-TOF HDMS and HPLC-DAD analysis. J Agric Food Chem 2010;58:4602-16.

[3] Yuan B, Byrnes D, Wu Q, Simon JE. Identification of polyphenols, glycoalkaloids and saponins in *Solanum scabrum* berries using HPLC-UV/Vis-MS J Food Sci 2018.

[4] R-Core-Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/. 2013.

[5] Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 2016;32:2847-9.

[6] Afari-Sefa V, Tenkouano A, Ojiewo CO, Keatinge JDH, d'A. Hughes J. Vegetable breeding in Africa: constraints, complexity and contributions toward achieving food and nutritional security. Food Security 2011;4:115-127.

[7] Defelice MS. The Black Nightshades, *Solanum nigrum* L. et al.—Poison, Poultice, and Pie. Weed Technol 2003;17:421-427.

[8] Matasyoh LG, Murigi HM, Matasyoh JC. Antimicrobial assay and phyto-chemical analysis of *Solanum nigrum* complex growing in Kenya. Afr J Microbiol Res 2014;8:3923-3930.

[9] Carle R. Investigations on the content of steroidal alkaloids and sapogenins within *Solanum* sect. *Solanum* (= sect. *Morella*)(Solanaceae). Plant Syst Evol 1981;138:61-71.

[10] Akula R, Ravishankar GA. Influence of abiotic stress signals on secondary metabolites in plants. Plant signaling & behavior 2011;6:1720-1731.

[11] Chalker-Scott L. Environmental significance of anthocyanins in plant stress responses. Photochemistry and photobiology 1999;70:1-9.

[12] Mathe Jr I, Van Mai H, Máthé I. Studies on the *Morella* section of the *Solanum* genus. V. Evaluation of the alkaloid production of *Solanum americanum* Mill. Acta agronomica 1980.

[13] Hutchinson A, Hilton R. The influence of certain cultural practices on solanine content and tuber yields in Netted Gem potatoes. Canadian Journal of Agricultural Science 1955;35:485-491.

[14] Yousuf B, Gul K, Wani AA, Singh P. Health Benefits of Anthocyanins and Their Encapsulation for Potential Use in Food Systems: A Review. Crit Rev Food Sci Nutr 2016;56:2223-30.

[15] Dai J, Mumper RJ. Plant phenolics: extraction, analysis and their antioxidant and anticancer properties. Molecules 2010;15:7313-52.

[16] Price CW, Ronald E. Anthocyanin pigments of Royal Okanogan huckleberry juice. J Food Sci 1995;60:369-374.

[17] Vollmannova A, Toth T, Urminska D, Polakova Z, Timoracka M, Margitanova E. Anthocyanins content in blueberries (*Vaccinium corymbosum* L.) in relation to freezing duration. Czech J Food Sci 2009;27:204-206.

[18] Prior RL, Cao G, Martin A, Sofic E, McEwen J, O'Brien C, Lischner N, Ehlenfeldt M, Kalt W, Krewer G, Mainland CM. Antioxidant Capacity As Influenced by Total Phenolic

and Anthocyanin Content, Maturity, and Variety ofVacciniumSpecies. Journal of Agricultural and Food Chemistry 1998;46:2686-2693. [In English]

[19] Francis F, Harborne J. Anthocyanins of the garden huckleberry, *Solanum guineese*. J Food Sci 1966;31:524-528.

[20] Fossen T, Cabrita L, Andersen OM. Colour and stability of pure anthocyanins influenced by pH including the alkaline region. Food Chem 1998;63:435-440.

[21] Patras A, Brunton NP, O'Donnell C, Tiwari BK. Effect of thermal processing on anthocyanin stability in foods; mechanisms and kinetics of degradation. Trends Food Sci Technol 2010;21:3-11.

[22] Joint FAO/WHO Expert Committee on Food Additives WHO, Food and Agriculture Organization of the United Nations. Evaluation of certain food additives and naturally occurring toxicants : thirty-ninth report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva, Switzerland. 1992.

[23] Bajaj K, Kaur G, Chadha M. Glycoalkaloid content and other chemical constituents of the fruits of some eggplant (*Solanum melongena*. L.) varieties. J Plant Foods 1979.

[24] Jones PG, Fenwick GR. The glycoalkaloid content of some edible solanaceous fruits and potato products. Journal of the Science of Food and Agriculture 1981;32:419-421.

[25] Friedman M, McDonald GM, Filadelfi-Keszi M. Potato Glycoalkaloids: Chemistry, Analysis, Safety, and Plant Physiology. Crit Rev Plant Sci 2010;16:55-132.

[26] Sim HJ, Yoon SH, Kim MS, Kim B, Park HM, Hong J. Identification of alkaloid constituents from *Fangchi* species using pH control liquid-liquid extraction and liquid chromatography coupled to quadrupole time-of-flight mass spectrometry. Rapid Commun Mass Spectrom 2015;29:837-54.

[27] Wrolstad RE, Rodriguez-Saona LE. Natural colorant from potato extract. Google Patents; 2001.

[28] Johns T. Detoxification function of geophagy and domestication of the potato. J Chem Ecol 1986;12:635-46.

[29] Weng A, Thakur, Melzig, Fuchs. Chemistry and pharmacology of saponins: special focus on cytotoxic properties. Botanics: Targets and Therapy 2011.

[30] Qin Y, Wu X, Huang W, Gong G, Li D, He Y, Zhao Y. Acute toxicity and sub-chronic toxicity of steroidal saponins from *Dioscorea zingiberensis* C.H.Wright in rodents. J Ethnopharmacol 2009;126:543-50.

[31] Zhang X, Jin M, Tadesse N, Xian L, Zhang H, Wang S, Dang J, Zhang Y, Guo Z, Ito Y. Safety investigation on total steroid saponins extracts from *Dioscorea zingiberensis* C.H. Wright: Sub-acute and chronic toxicity studies on dogs. Regul Toxicol Pharmacol 2017;91:58-67.

[32] Xu T, Zhang S, Zheng L, Yin L, Xu L, Peng J. A 90-day subchronic toxicological assessment of dioscin, a natural steroid saponin, in Sprague-Dawley rats. Food Chem Toxicol 2012;50:1279-87.

[33] Nair AB, Jacob S. A simple practice guide for dose conversion between animals and human. J Basic Clin Pharm 2016;7:27-31.

[34] Yuan B, Byrnes D, Giurleo D, Villani T, Simon JE, Wu Q. Rapid screening of toxic glycoalkaloids and micronutrients in edible nightshades (*Solanum* spp.). J Food Drug Anal 2018;26:751-760.

CHAPTER VIII. ENHANCED METHOD FOR BERRY QUALITY CONTROL

1 Introduction

This work proposed a new method for glycoalkaloid screening with significantly enhanced throughput and smoother workflow. Distinct from the prior untargeted profiling in CHAPTER VII, this work developed a targeted analysis methodology specific to glycoalkaloids using triple quadruple mass spectrometry featuring pseudo-MS/MS/MS. In particular, in-source fragmentation (ISF) was innovatively applied as a pseudo-MS or pseudo-hydrolysis to break down glycoalkaloidal glycosides into corresponding aglycones prior to MS/MS detection. This approach successfully overcame the difficulty of long cycle time and slow scan speed of traditional MS/MS when addressing large numbers of glycosidic compounds, and made the targeted methodology ever feasible in this work. The new method developed could meet the demand of the most heavy-duty glycoalkaloids analysis in a fast and easy-to-use workflow with least manual intervention.

2 Materials and Methods

2.1 Chemicals

Standard reference compound solamargine as the only commercially available standard of SNB glycoalkaloids was purchased from MedChemExpress (NJ, USA), LC/MS grade reagents including formic acid, acetonitrile and methanol from Fisher Scientific (Fair Lawn, NJ, USA) and water from Thermo Fisher Scientific (Waltham, MA, USA).

2.2 Plant material

Berries of two *S. scabrum* accessions, Ex Hai and BG 29 from World Vegetable Center (WorldVeg, Arusha, Tanzania) and two *S. nigrum* accessions, one from a private seed company Simlaw Seeds (Nairobi City, Kenya) and the other PI 381289 from U. S. Department of Agriculture (USDA, Ames, IA, USA), were analyzed in this work. The samples originated from the same field trail as reported in the prior work [1, 2]. Briefly, the seeds were sown in 72-cell trays with growing mix (Fafard Grow Mix 2; Sun Gro Horticulture, Agawam, MA) and germinated at the New Jersey Agricultural Experiment Station (NJAES) Research Greenhouse of Rutgers University. The seedlings were transplanted after four weeks of growth to raised beds covered with 1.25mm black plastic mulch with drip irrigation applied as needed in NJAES Horticultural Research Farm #3, New Brunswick, New Jersey. The mature berries, with deep purple color in both skin and flesh, were harvested on September 18, 2016 and kept frozen under -20 °C, and thawed under room temperature upon analysis.

2.3 Instrument

The instrument used for chemical analysis was Agilent 1290 Infinity II UHPLC hyphenated with 6470 triple quadrupole mass spectrometry with electrospray ionization source (ESI) (Santa Clara, CA, USA) [3]. MassHunter Workstation software Data Acquisition (version B.08.00) and Quantitative Analysis (version B.07.01) were used for data processing. Chromatographic separation of compounds was achieved using Waters

Acquity UHPLC BEH C18 column (50 × 2.1 mm, 1.7 μm) equipped with Acquity BEH C18 guard column (5 × 2.1 mm, 1.7 μm) (Milford, MA, USA).

2.4 Sample preparation

Around 6 g of berries were grounded and soaked in 40 mL 70% aqueous methanol containing 0.1% formic acid, vigorously vortexed for 1 min and sonicated for 15 minutes. An aliquot of 30 μL of the extract was diluted by mixing with 0.9 mL 70% methanol with 0.1% formic acid, centrifuged at 16,000 rpm for 10 minutes, and then the supernatant was injected for LC-MS analysis.

2.5 LC-MS method development

For LC condition, water with 0.1% formic acid was used as mobile phase A and acetonitrile with 0.1% formic acid as mobile phase B. The flow rate was 0.4 mL/min. The gradient started from 15% B and increased to 31% B at 5min, 31% to 35% B from 5 to 5.1 min, 35% to 55% B from 5.1 to 6min, and was isocratically kept at 55% B from 6 to 6.5min, followed by column equilibration with the starting mobile phase for another 0.5 min. Eluent from 0 to 0.5 min was diverted to waste. The column thermostat was set at 30 °C. The autosampler was maintained at 4 °C, and the injection volume was 1 μL. A 3-second needle wash was applied using 70% methanol after each sample injection.

For MS conditions, the drying gas was set at 250°C at 13 L/min, sheath gas temperature was 300°C at 12 L/min. The nebulizer pressure was 30 psi. Positive polarity was used, with nozzle voltage at 1000 V and capillary voltage at 3500 V. Multiple reaction monitoring (MRM) was applied as scanning mode, with dwell time 20 ms per transition.

2.6 Method validation

The method developed was validated in terms of sensitivity, calibration range, accuracy, matrix effect and precision for compound solamargine, which is the only SNB glycoalkaloid whose reference standard is commercially available. Briefly, the lower limit of detection (LLOD) is defined as the injection concentration to give signal-to-noise ratio (S/N) of 3 and lower limit of quantification (LLOQ) with S/N at 10. For accuracy validation, aliquots of 1 mL berry extract from accession WorldVeg BG 29 prepared as mentioned above were used as the quality control sample (QCS), and were spiked with 70, 135 and 270 μL of 9 μg/mL solamargine stock solution, which corresponded to *ca* 50, 100 and 200 % of the solamargine concentration in the original QCS, respectively, with each spike level prepared in three replicates. The samples were diluted by 31 times and then centrifuged prior to LC/MS analysis. The accuracy was computed as the measured concentration in spiked QCS subtracting that in original QCS then divided by the expected spiked level. For matrix validation, aliquots of 1 mL pure solvent of 70% methanol were spiked with solamargine standard solution in like manner as accuracy validation, and the matrix effect was computed as the measured concentration in spiked QCS subtracting that in original QCS then divided by the measured concentration in pure solvents. The standard deviation of accuracy and matrix effect is computed following the law of error propagation, with formulas shown in the supplementary material. In addition, the standard deviation of concentrations in spiked pure solvents measured at each spike level of the same set used in matrix validation is reported as the intra-batch precision.

2.7 Statistics

ESI conditions were optimized using central composite design (CCD) [4] (**Table A-6**). Data analysis and visualization was achieved using R [5], and the script is shown in https://yuanbofaith.github.io/Solanum_alkaloid_in-source-fragmentation_MSMS/.

3 Results and Discussion

3.1 LC-MS method development

3.1.1 Precursor ion selection and in-source fragmentation

While *intact parent ions* in protonated or other adducted forms are commonly selected as MRM precursors, it does not conveniently apply to SNB glycoalkaloids in this work. Since glycoalkaloids as glycosidic compounds have very flexible combinatorial structures composed of different aglycones and oligosaccharide chains, they present a wide range of possible structures of varied masses. Trying to enumerate all theoretically possible combinatorial structures and precursors to create a complete MRM transition table would be daunting, which when constructed would also result in long cycle time and slow scan speed, and therefore compromise sensitivity and limit the number of collectable data points across a chromatographic peak, rendering the peak unsmooth and even incomplete [6] (also see **Figure A-17** for association between the number of collectable data points with chromatographic peak width, MRM dwell time, cycle time and transition number). The long cycle time could be easily decreased by using shorter dwell time which, however, could be at the sacrifice of method sensitivity. Updating the MRM method to the dynamic version (dMRM) may help increase allotted dwell time for each transition and/or increase collectable data points across chromatographic peaks [7], but simultaneous determination

of the retention time for all glycoalkaloids listed in the transition table in this specific work is very difficult to successfully achieve. Selection of intact glycoalkaloids as the precursor ions further complicates providing a clear solution.

However, this dilemma was innovatively resolved by use of ISF of glycoalkaloids. For the "source" of an MS, it typically refers to the electrospray ionization (ESI) chamber, the transmission capillary and other spaces maintained under intermediate vacuum by the rough pump prior to the internal high vacuum. While ESI is generally considered as a "soft" ionization technique, ions with weak bonds can still undergo fragmentation during ionization and desorption from the aerosol in the ESI chamber, with fragmentation intensity affected by the mobile phase and the gas and voltage settings of the ESI [8-10]. At the end of the transmission capillary, fragmentor voltage is applied to propel ion transmission, and additional fragmentation could happen due to collision between accelerated ions with the surrounding neutral gas (this is also known as the in-source collision-induced dissociation (CID), which is distinguished from the CID occurring in the designated collision cell by clash with auxiliary gas like nitrogen, helium or argon) [11, 12]. As such, fragmentation intensity could also be affected by fragmentor voltage. While ISF has usually been considered as an undesirable phenomenon which adds complexity and confusion to MS analysis [8, 11, 13], it has also been drawn upon for specialized analysis. For example, ISF was applied for gain of structural information without using MS/MS technique [14]; as replacement of traditional hydrolysis to break down singly-charged polysaccharides into measurable pieces within the limit of instrument scanning range [15]; or to distinguish coeluted isobaric compounds [16]. In this work, ISF was applied as a *pseudo*-hydrolysis to break down the glycoalkaloids' glycosidic bond to release the aglycone ions, which were

then subjected to MS/MS analysis, collectively constituting a *pseudo*-MS/MS/MS analysis. Therefore, ISF trimmed the complexity of the rich diversity of glycosides down to simplicity of five basic types of aglycones (including isomers), i.e., solasodine (S) at 414 *m*/*z*, hydroxysolasodine (HS) 430 *m*/*z*, dihydroxysolasodine (DHS) 446 *m*/*z*, hydroxymethylsolasodine (HMS) 444 *m*/*z*, and dihydroxymethylsolasodine (DHMS) 460 *m*/*z* (see structure in prior Chapter **Figure VI-1**). The mechanism of ISF-MS/MS or *pseudo*-MS^3 is further illustrated in **Figure VIII-1**. As such, transitions only related to the five precursor masses were needed to construct the MRM table, which significantly reduced the total cycle time and improved scan speed. Meanwhile, unlike the glycoside-mediated MRMs that are "blind" to glycosides not included in the fixed MRM table, the aglycone-mediated MRM is barely limited by unconsidered glycosylation pattern, such as introduction of new saccharides or change in the size of the oligosaccharide chain, making the analysis targeted to but also universal in glycoalkaloid detection.

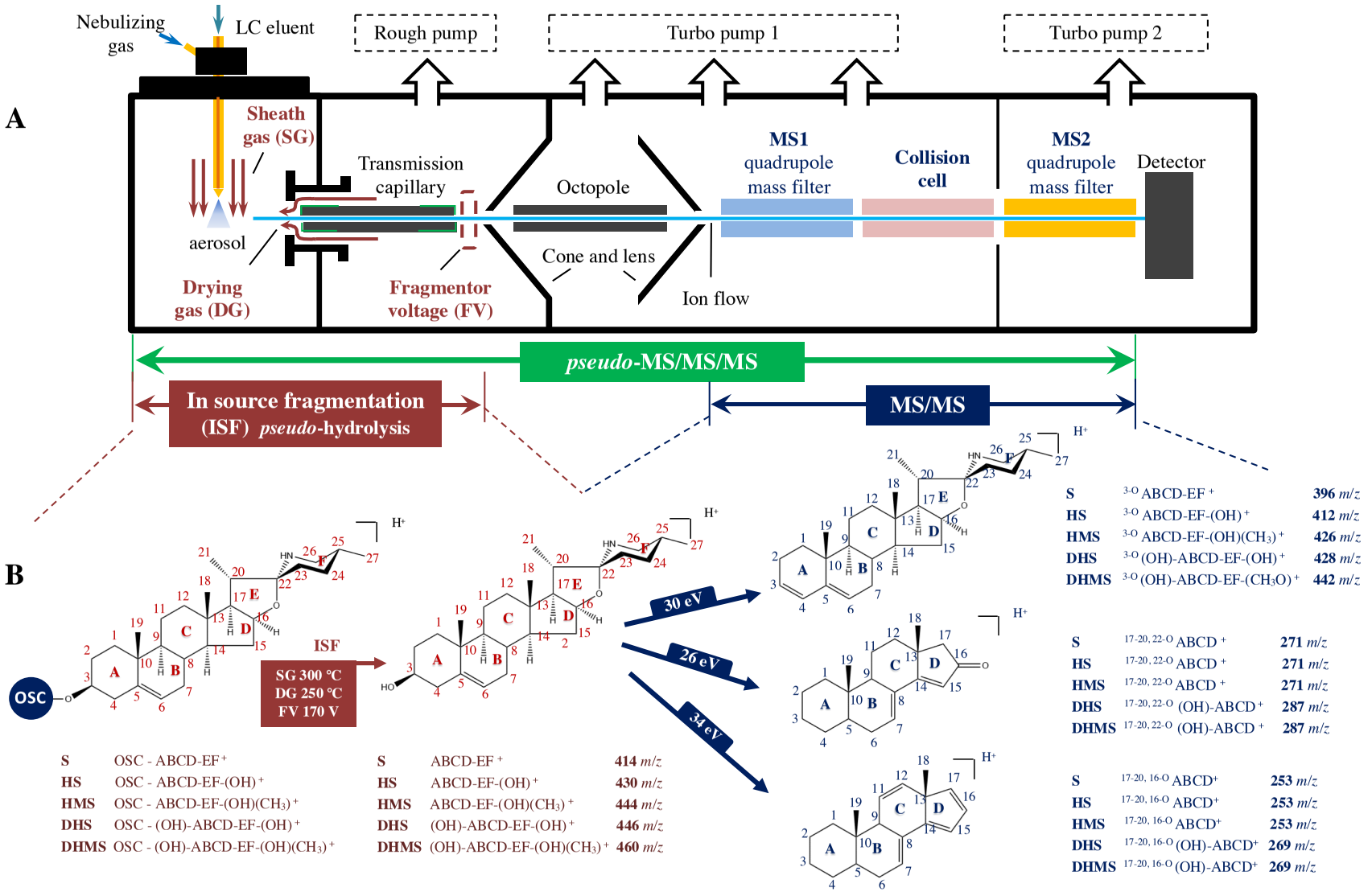

A
Nebulizing gas
LC eluent
Rough pump
Turbo pump 1
Turbo pump 2
Sheath gas (SG)
Transmission capillary
Octopole
MS1 quadrupole mass filter
Collision cell
MS2 quadrupole mass filter
Detector
aerosol
Drying gas (DG)
Fragmentor voltage (FV)
Cone and lens
Ion flow
pseudo-MS/MS/MS
In source fragmentation (ISF) pseudo-hydrolysis
MS/MS
B
OSC
ISF
SG 300 °C
DG 250 °C
FV 170 V
30 eV
26 eV
34 eV
S OSC - ABCD-EF+
HS OSC - ABCD-EF-(OH)+
HMS OSC - ABCD-EF-(OH)(CH3)+
DHS OSC - (OH)-ABCD-EF-(OH)+
DHMS OSC - (OH)-ABCD-EF-(OH)(CH3)+
S ABCD-EF+ 414 m/z
HS ABCD-EF-(OH)+ 430 m/z
HMS ABCD-EF-(OH)(CH3)+ 444 m/z
DHS (OH)-ABCD-EF-(OH)+ 446 m/z
DHMS (OH)-ABCD-EF-(OH)(CH3)+ 460 m/z
S 3-O ABCD-EF+ 396 m/z
HS 3-O ABCD-EF-(OH)+ 412 m/z
HMS 3-O ABCD-EF-(OH)(CH3)+ 426 m/z
DHS 3-O (OH)-ABCD-EF-(OH)+ 428 m/z
DHMS 3-O (OH)-ABCD-EF-(CH3O)+ 442 m/z
S 17-20, 22-O ABCD+ 271 m/z
HS 17-20, 22-O ABCD+ 271 m/z
HMS 17-20, 22-O ABCD+ 271 m/z
DHS 17-20, 22-O (OH)-ABCD+ 287 m/z
DHMS 17-20, 22-O (OH)-ABCD+ 287 m/z
S 17-20, 16-O ABCD+ 253 m/z
HS 17-20, 16-O ABCD+ 253 m/z
HMS 17-20, 16-O ABCD+ 253 m/z
DHS 17-20, 16-O (OH)-ABCD+ 269 m/z
DHMS 17-20, 16-O (OH)-ABCD+ 269 m/z

Figure VIII-1. In-source fragmentation (ISF) and tandem mass spectrometry (MS/MS) analysis procedure of glycoalkaloids. (**A**), instrument schematic. Note that for instrument used in this work, the collision cell assumes a curvature configuration (shown in linear instead for simplicity). (**B**), glycoalkaloids and associated precursor and fragmental ions along the procedure of ISF-MS/MS. For fragmental ion nomenclature, the bonds ruptured are noted as the hyphenated carbon numbers or adjacent heteroatomic symbols, marked as superscripts of ring notations A to F. The fused ABCD rings and the spiro EF ring are arbitrarily noted as two sections, respectively, with the ABCD ring likely substituted with hydroxyl group, and the EF ring likely substituted with hydroxyl and methyl or isomeric substituents. In (**A**) and (**B**), instrument components and molecular ions associated with ISF are noted in red, and those associated with MS/MS in blue. ISF and MS/MS together constitutes *pseudo*-MS^3. For abbreviations of instrumental components and parameters, DG, drying gas; SG, sheath gas, FV, fragmentor voltage; CID, collision-induced dissociation in the collision cell. Compound abbreviations refer to prior chapter **Table VI-1**.

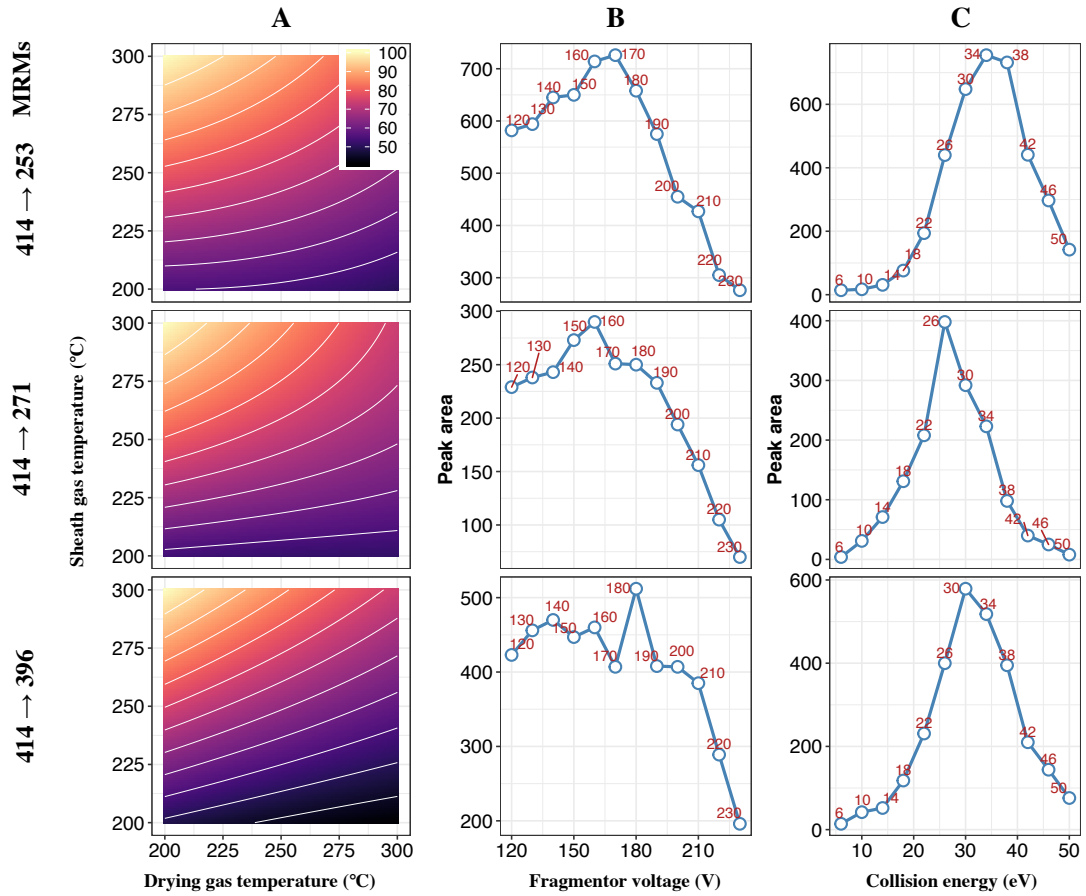


Figure VIII-2. Optimization of instrument parameters. (**A**) contour plot of signal response vs. sheath gas temperature and drying gas temperature; (**B**) the signal response vs. fragmentor voltage; (**C**) response vs. collision energy. Solamargine standard was used for optimization as representative glycoalkaloid in *S. nigrum* and *scabrum* berries. Effect of drying gas temperature and sheath gas temperature was modeled and optimized using central composite design (CCD) to achieve maximum yield of aglycone solasodine 414 *m/z* generated by in-source fragmentation, with its abundance measured by corresponding product ions.

To achieve maximum yield of aglycones via in-source fragmentation, key instrumental parameters were optimized. A CCD experiment was conducted to optimize the ESI drying gas temperature and sheath gas temperature using solamargine standard as the representative glycoalkaloid, with result shown in **Figure VIII-2A** (see above). Interestingly, the temperature of the sheath gas more than that of the drying gas was found to be a more critical factor to "overcook" ions for fragmentation. Suggested by the less

curly contours, sheath gas and drying gas did not present as much interaction or complementary effect as expected and observed in a prior study [9]. Based on the CCD results, the drying gas temperature was set at 250°C and sheath gas temperature was 300°C. In addition, the fragmentor voltage as aforementioned was also optimized, leading to an optimal fragmentor voltage at 170 V on average as shown in **Figure VIII-2B** (see above).

3.1.2 Product ion selection and fragmentation pathway

The characteristic ions of glycoalkaloidal aglycones were selected based on structural scaffold, substitution pattern and corresponding fragmentation pathway (**Table VIII-1** and **Figure A-16**) [1, 17, 18]. Since different glycoalkaloidal aglycones share similar structural skeleton, then the calibration curve constructed from solamargine could be used to estimate the concentration of other glycoalkaloids detected via MRMs of the same fragmentation pathway (so termed the "principle of same fragmentation pathway" for convenience of discussion). This is a convenient way to detect and also estimate the quantity of the glycoalkaloids in absence of the corresponding authenticated standards.

For association between the product ions with aglycone substitution pattern and fragmentation pathway, the precursor ion of solasodine 414 *m/z*, i.e., the aglycone of solamargine, featured three most abundant and characteristic product ions, $^{3\text{-O}}ABCDEF^{+}$ of 396 *m/z* by losing the 3-hydroxyl group from the aglycone precursor ion, $^{17\text{-}20,\,22\text{-O}}ABCD^{+}$ of 271 *m/z* by the following rupture of the E ring, and $^{17\text{-}20,\,16\text{-O}}ABCD^{+}$ of 253 *m/z* by subsequent loss of the carbonyl group. Compared with the solasodine precursor, the HS aglycone has one more hydroxyl group on the E-F ring; DHS has one more hydroxyl group respectively on the E-F ring and A-B-C-D ring; HMS one additional hydroxyl and methyl

(or a combined form as methoxyl or other equivalent forms) on the E-F ring; and DHMS one additional hydroxyl on the A-B-C-D ring, and additional hydroxyl and methyl on the E-F ring as the case of HMS. As such, the masses of product ions from HS, DHS, HMS and DHMS along the proceedings of the fragmentation pathway would be higher by the mass of corresponding substitution relative to those of solasodine**.**

In order to achieve maximum abundance from each transition, the MRM collision energy was optimized using solamargine as reference standard, as shown in **Figure VIII-2C**. The product of 253 *m/z*, generated by collision energy (CE) 34 eV, was of the most abundance and used as the quantifier for calibration construction; 271 *m/z* with CE 26 eV, and 396 *m/z* with CE 30 eV were used as the qualifier ion. For all other glycoalkaloids, the transitions were setup based on aglycone substitution and fragmentation pathway relative to that of solamargine as discussed above.

3.2 LC-MS method validation

As solamargine was one of the most abundant and characteristic glycoalkaloids in SNB, and that its authentic standard was also available, the quantification of this compound was validated, with results shown in **Table VIII-1**. Validated at three spike levels corresponding to 50, 100 and 200% of the solamargine content in the original berry extract, the validated accuracy was 77~79%, matrix effect from the 84~95% and precision 0.8~4%. The linearity range had three orders of magnitude with low limit of quantification down to 1 ng/mL or 1 pg injected on column. Since the standards of other glycoalkaloids are not available, a quantitative validation of these compounds was not feasible.

Table VIII-1. Multiple reaction monitoring (MRM) for glycoalkaloids in *Solanum scabrum* and *S. nigrum* berries, and associated validation results of solamargine.

Glycoalkaloid series	Precursor ion (*m/z*)	Fragmentor voltage (V)	Product ion (*m/z*)	Collision energy (eV)
DHMS glycosides (460 *m/z* series)	460	170	442	30
	460	170	287	26
	460	170	269	34
DHS glycosides (446 *m/z* series)	446	170	428	30
	446	170	287	26
	446	170	269	34
HMS glycosides (444 *m/z* series)	444	170	426	30
	444	170	271	26
	444	170	253	34
HS glycosides (430 *m/z* series)	430	170	412	30
	430	170	271	26
	430	170	253	34
S glycosides (414 *m/z* series)	414	170	396	30
	414	170	271	26
	414	170	253	34

LLOD (ng/ml)	LLOQ (ng/ml)	Linear range (ng/ml)	Calibration curve	R^2
0.07	1.10	1.10 ~ 1125	Y=19.1561 X + 24.9620	0.9939 ‡

At different spike levels: Accuracy (%)			Matrix effect (%)			Precision (%)		
50%	100%	200%	50%	100%	200%	50 %	100 %	200%
78.2 ± 4.5	78.9 ± 2.9	77.7 ± 1.0	84.4 ± 2.3	95.0 ± 4.2	91.4 ± 1.1	0.6	3.9	0.8

For compound abbreviations, S, solasodine; HS, hydroxysolasodine or isomers; HMS, hydroxymethylsolasodine, methoxysolasodine or other possible isomers; DHS, dihydroxysolasodine and isomers; DHMS, dihydroxymethylsolasodine, hydroxymethoxysolasodine or other isomers. For the validation table, ‡, R^2 was calculated with 1/x weight. Accuracy, matrix effect and precision were validated at three spike levels corresponding to *ca.* 50, 100 and 200% of the solamargine content in the original berry extract.

3.3 Glycoalkaloids in SNB

The glycoalkaloids from four different SNB were successfully detected using the developed method featuring ISF-MS/MS. A typical chromatogram of glycoalkaloids in a SNB sample is shown in **Figure VIII-3**. Glycosides from each type of aglycone, as called the series of DHMS, DHS, HMS, HS and S, were eluted out generally with increasing retention time due to reduced polarity of the aglycones. Each aglycone-based series were conveniently detected in three transition windows, one quantifier transition and two qualifier ones in such targeted manner, and identified simply by associating detected peaks with corresponding characteristic transitions, free from mass spectra interpretation. In addition, based on the principle of the same fragmentation pathway aforementioned, the glycoalkaloidal content in the SNB samples were also determined, shown in **Table VIII-2** (note that in CHAPTER VII, quantification was conducted based on dehydrated berries, while this work on frozen fresh berries). In all four sourced fresh berries, the total *aglycone* mass ranged from trace content to 1.8 mg/100g fresh weight (FW), or approximately total *glycosides* around 4 mg/100g FW. When compared with the reported safe content of solamargine and analogous compounds existing in eggplant (*Solanum melongena* L.) ranging from 6.25 to 20.5 mg/100 g FW [19, 20], and also referenced with the safety threshold of potato (*Solanum tuberosum* L.) glycoalkaloids at 10 mg/100g FW [21], the content of glycoalkaloids in the analyzed samples was lower and might not pose a noticeable health hazard at least when consumed with small amount. In addition, the *S. nigrum* accession from USDA PI 381289 contained only trace amount of glycoalkaloids (also refer to prior chapter **Table VII-2** for additional discussion in phytochemical profile), and could be recommended as a genetic line to serve as a safe source of fresh berries.

It is important to note that this quantitative determination still remains approximate in nature. This is best demonstrated by the different ratio of product ions' abundances vs. those of solamargine which was used to construct the calibration curve. The product ion of 253 *m/z* from solamargine, for example, displayed the highest abundance, but for the first compound of the HS series showed far less abundance than other product ions (**Figure VIII-3**). In fact, the HS aglycones likely presented more than one possible structure manifested by the multifold of chromatographically resolved aglycones after hydrolysis [1], either possibly different in the hydroxyl substitution position or in their stereo isomerization, and all such difference could make a difference in the abundance ratio of the final product ions. In addition, the principle of the same fragmentation pathway could also be challenged by differences of glycoalkaloids' efficiency in ionization and ISF. As such, sufficient safety margin should be given when decisions are made about berry toxicity based on the quantity-estimated glycoalkaloidal profile.

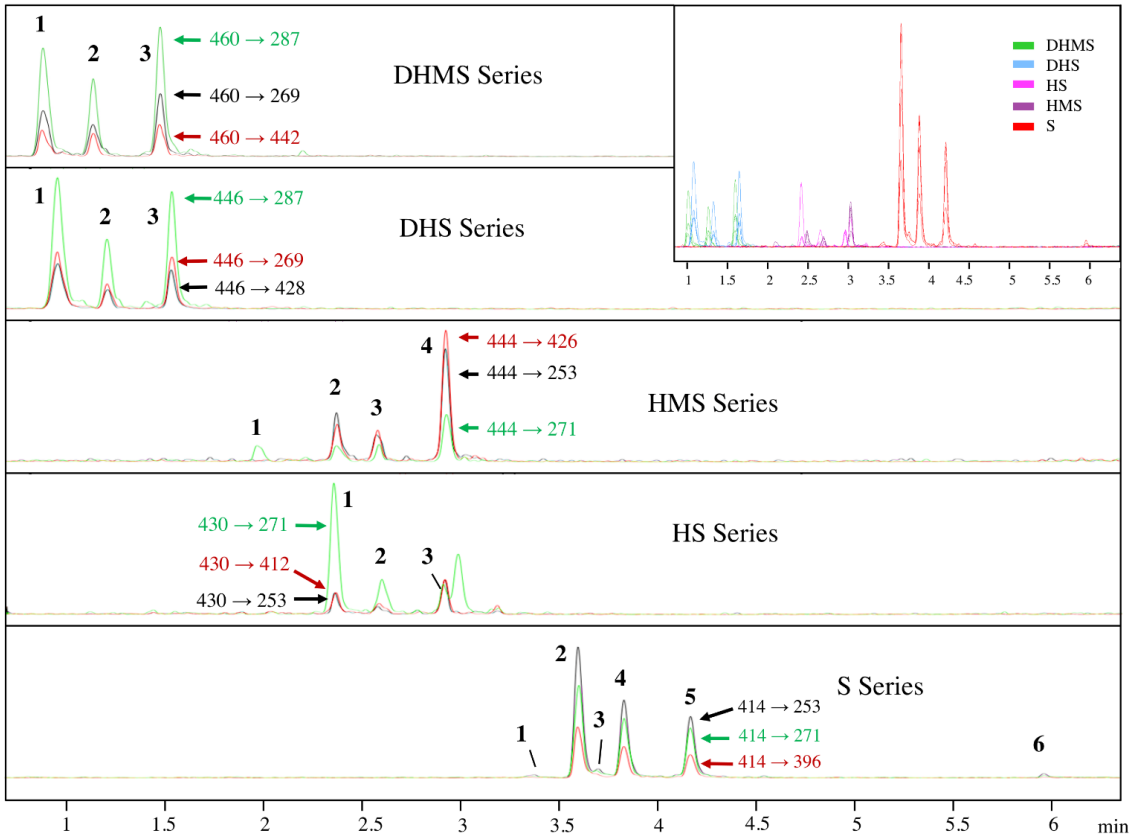


Figure VIII-3. Representative chromatograms obtained with in-source fragmentation (ISF) and multiple reaction monitoring (MRM), with ISF-generated aglycones selected as precursor ions of MRM transitions. The data shown was acquired from *Solanum scabrum* berries of BG 29 from World Vegetable Center (Arusha, Tanzania). Compound abbreviations refer to prior chapter **Table VI-1**. Transitions from the same step along the proceedings of fragmentation pathway are noted as the same color. Compound **6** was identified as free aglycone of solasodine by comparison with authentic standard.

Table VIII-2. Content of glycoalkaloids in fresh berries of *Solanum scabrum* and *Solanum nigrum*.

Compounds	Code	RT (min)	*S. scabrum* WorldVeg Ex Hai		*S. scabrum* WorldVeg BG-29		*S. nigrum* USDA PI 381289		*S. nigrum* Simlaw Seeds	
			A	B	A	B	A	B	A	B
DHMS glycosides (460 *m/z* series)	**1**	0.73	0.13	0.06	0.16	0.07	N.D.	N.D.	0.14	0.06
	2	1.00	0.10	0.04	0.11	0.05	N.D.	N.D.	0.01	0.01
	3	1.35	0.28	0.13	0.19	0.09	N.D.	N.D.	0.17	0.08
	subtotal		***0.50***	***0.23***	***0.46***	***0.21***	***N.D.***	***N.D.***	***0.32***	***0.15***
DHS glycosides (446 *m/z* series)	**1**	0.81	0.01	0.01	0.23	0.10	N.D.	N.D.	1.08	0.48
	2	1.07	T	T	0.05	0.02	N.D.	N.D.	0.10	0.05
	3	1.42	0.01	0.01	0.13	0.06	N.D.	N.D.	0.46	0.21
	4	2.67	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	0.07	0.03
	subtotal		***0.03***	***0.01***	***0.42***	***0.19***	***N.D.***	***N.D.***	***1.71***	***0.76***
HMS glycosides (444 *m/z* series)	**1**	1.88	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	0.02	0.01
	2	2.31	N.D.	N.D.	0.07	0.03	N.D.	N.D.	0.00	0.00
	3	2.54	T	T	0.04	0.02	N.D.	N.D.	T	T
	4	2.89	0.02	0.01	0.21	0.09	N.D.	N.D.	0.02	0.01
	subtotal		***0.02***	***0.01***	***0.32***	***0.14***	***N.D.***	***N.D.***	***0.05***	***0.02***
HS glycosides (430 *m/z* series)	**1**	2.23	N.D.	N.D.	0.02	0.01	N.D.	N.D.	T	T
	2	2.47	N.D.	N.D.	T	T	N.D.	N.D.	N.D.	N.D.
	3	2.82	T	T	0.06	0.02	N.D.	N.D.	T	T
	subtotal		***T***	***T***	***0.08***	***0.03***	***N.D.***	***N.D.***	***T***	***T***
S Series (414 *m/z* series)	**1**	3.32	N.D.	N.D.	0.02	0.01	N.D.	N.D.	0.12	0.05
	2	3.55	0.04	0.01	1.41	0.58	N.D.	N.D.	0.04	0.02
	3	3.66	T	T	0.09	0.04	N.D.	N.D.	0.13	0.05
	4	3.79	0.13	0.06	0.83	0.34	T	T	0.01	0.00
	5	4.15	0.10	0.04	0.67	0.28	N.D.	N.D.	0.00	0.00
	6	6.01	0.00	0.00	0.01	0.00	T	T	0.00	0.00
	subtotal		***0.27***	***0.11***	***3.03***	***1.25***	***T***	***T***	***0.30***	***0.12***
Total alkaloid			***0.81***	***0.36***	***4.24***	***1.80***	***T***	***T***	***2.38***	***1.05***

The content is reported respectively in unit of μmol glycoalkaloid / 100 g fresh weight (FW) in column A, and mg aglycone / 100 g FW in column B. Aglycone abbreviations refer to the prior chapter **Table VI-1**, and individual compound codes correspond to **Figure VIII-3**. The plants are noted as species identity, source and accession number if applicable. Berry germplasms were collected from World Vegetable Center (WorldVeg, Arusha, Tanzania); U.S. Department of Agriculture (USDA, Ames, IA, USA), and a private seed company Simlaw Seeds (Nairobi City, Kenya). For content notation, T, trace amount; N.D., not detected.

4 Conclusion

This work developed a rapid detection method for SNB glycoalkaloids using UHPLC-ISF-MS/MS. The innovative use of ISF successfully overcame the difficulty posed by the diversity and complexity of glycosides and the scarcity of reference standards, and allowed for application of MRM-based targeted analysis methodology to achieve the highest throughput. By selecting characteristic product ions that are generated from the same fragmentation pathway of that of solamargine, compound content estimation was also made possible, allowing for estimation of berry toxicity based on glycoalkaloidal profile. The method established was superior to prior untargeted profiling methodology by being glycoalkaloid-specific, free from interference of the rich existence of non-glycoalkaloidal compounds and the cumbersomeness of mass spectra interpretation and database search. It is a powerful tool for SNB glycoalkaloids quality control, safety inspection and breeding selection, etc. Part of the ongoing and future work involves the isolation of additional SNB glycoalkaloids as standards to use for more accurate quantification, and to investigate the association between glycoalkaloidal structures and signal responses relative to that of solamargine, which is the only one whose standard is commercially available, so as to provide the correction rules for glycoalkaloids quantification using solamargine as the sole standard.

5 References

[1] Yuan B, Byrnes D, Dinssa FF, Simon JE, Wu Q. Identification of Polyphenols, Glycoalkaloids, and Saponins in *Solanum scabrum* Berries Using HPLC-UV/Vis-MS. J Food Sci 2019;84:235-243.

[2] Yuan B, Byrnes D, Dinssa FF, Simon JE, Wu Q. Quantity assessment of polyphenols, glycoalkaloids and saponins in *Solanum scabrum* berries of different genetic sources and maturity by HPLC/UV-visible/MS methods. J Sci Food Agric 2019;99:3578-3587.

[3] Agilent-Technologies. Agilent 6400 Series Triple Quadrupole LC/MS System. Concepts Guide. Revision A. August 2015;Manual part number G3335-90205.

[4] Montgomery DC. Design and analysis of experiments: John wiley & sons; 2017.

[5] R-Core-Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/. 2013.

[6] Yuan B, Dinssa FF, Simon JE, Wu Q. Simultaneous quantification of polyphenols, glycoalkaloids and saponins in African nightshade leaves using ultra-high-performance liquid chromatography-tandem mass spectrometry with acid assisted hydrolysis and multivariate analysis. Food Chem 2019:126030.

[7] Zhao D, Yuan B, Carry E, Pasinetti GM, Ho L, Faith J, Mogno I, Simon J, Wu Q. Development and validation of an ultra-high performance liquid chromatography/triple quadrupole mass spectrometry method for analyzing microbial-derived grape polyphenol metabolites. J Chromatogr B Analyt Technol Biomed Life Sci 2018;1099:34-45.

[8] Zhang CX, Wang XY, Lin ZZ, Wang HD, Qian YX, Li WW, Yang WZ, Guo DA. Highly selective monitoring of in-source fragmentation sapogenin product ions in positive mode enabling group-target ginsenosides profiling and simultaneous identification of seven Panax herbal medicines. J Chromatogr A 2020:460850.

[9] Yuan B, Zhao D, Du R, Kshatriya D, Bello NT, Simon JE, Wu Q. A highly sensitive ultra-high performance liquid chromatography/tandem mass spectrometry method with in-source fragmentation for rapid quantification of raspberry ketone. J Food Drug Anal 2019;27:778-785.

[10] Yuan B, Zhao D, Kshatriya D, Bello NT, Simon JE, Wu Q. UHPLC-MS/MS method development and validation with statistical analysis: determination of raspberry ketone metabolites in mice plasma and brain Anal Chem 2020;(under review).

[11] Criscuolo A, Zeller M, Fedorova M. Evaluation of Lipid In-Source Fragmentation on Different Orbitrap-based Mass Spectrometers. Journal of the American Society for Mass Spectrometry 2019;31:463-466.

[12] Abranko L, Garcia-Reyes JF, Molina-Diaz A. In-source fragmentation and accurate mass analysis of multiclass flavonoid conjugates by electrospray ionization time-of-flight mass spectrometry. J Mass Spectrom 2011;46:478-88.

[13] Xu YF, Lu W, Rabinowitz JD. Avoiding misannotation of in-source fragmentation products as cellular metabolites in liquid chromatography-mass spectrometry-based metabolomics. Anal Chem 2015;87:2273-81.

[14] Pais P, Moyano E, Puignou L, Galceran MT. Liquid chromatography-electrospray mass spectrometry with in-source fragmentation for the identification and quantification of fourteen mutagenic amines in beef extracts. J Chromatogr A 1997;775:125-36.

[15] Li J, Wang Z, Altman E. In-source fragmentation and analysis of polysaccharides by capillary electrophoresis/mass spectrometry. Rapid Commun Mass Spectrom 2005;19:1305-14.

[16] Carrier DJ, Eckers C, Wolff JC. "In-source" fragmentation of an isobaric impurity of lamotrigine for its measurement by liquid chromatography tandem mass spectrometry after pre-concentration using solid phase extraction. J Pharm Biomed Anal 2008;47:731-7.

[17] Cahill MG, Caprioli G, Vittori S, James KJ. Elucidation of the mass fragmentation pathways of potato glycoalkaloids and aglycons using Orbitrap mass spectrometry. J Mass Spectrom 2010;45:1019-25.

[18] Lelario F, Labella C, Napolitano G, Scrano L, Bufo SA. Fragmentation study of major spirosolane-type glycoalkaloids by collision-induced dissociation linear ion trap and infrared multiphoton dissociation Fourier transform ion cyclotron resonance mass spectrometry. Rapid Commun Mass Spectrom 2016;30:2395-2406.

[19] Bajaj K, Kaur G, Chadha M. Glycoalkaloid content and other chemical constituents of the fruits of some egg plant (*Solanum melongena*, L.) varieties. J Plant Foods 1979;3:163-168.

[20] Jones PG, Fenwick GR. The glycoalkaloid content of some edible solanaceous fruits and potato products. Journal of the Science of Food and Agriculture 1981;32:419-421.

[21] Joint FAO/WHO Expert Committee on Food Additives WHO, Food and Agriculture Organization of the United Nations. Evaluation of certain food additives and naturally occurring toxicants : thirty-ninth report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva, Switzerland. 1992.

CHAPTER IX. SUMMARY AND FUTURE WORK

Given the nutritional richness of the edible African nightshades (*Solanum* spp., primarily *S. scabrum* and *S. nigrum*, etc.) and the absence of anti-nutritive factors, specifically the glycoalkaloids in the leaves of these vegetables, the undervalued indigenous or traditional leafy greens could be playing far more important roles in enhancing food security, improving nutrition, and contributing toward agricultural and economic development in sub-Saharan Africa. This work that focuses on African nightshades, just one of the many African indigenous vegetables (AIVs), contributes new knowledge and insight into the chemistry of these vegetables, and provides the needed scientific guidance for safe and nutritious consumption. In summary, various LC-MS methods were developed for phytochemical determination to support exploratory analysis and/or high-throughput quality control. Using developed methods, the leafy phytochemistry in different accessions of nightshades were analyzed and quantified. The profile showed a deficiency of toxic glycoalkaloids, substantiating the safe consumption status of the leaves from all nightshades investigated. The profile of free amino acids was determined in leaves of an expanding number of AIVs, including nightshades, amaranth, mustard and spider plant. The total free amino acids presented high content ranging from several to ten percent of the dry mass. Machine learning methods and an interactive online program were applied and constructed for AIV classification prediction based on the profile of free amino acids.

Apart from leafy chemistry, the berry bioactive compounds in nightshades were also quantitfied. A range of glycoalkaloids with potentially novel structure were discovered with the structure tentatively identified. Content of glycoalkaloids differed significantly in

different genetic sources. Certain genetic lines were found lacking in glycoalkaloids while meantime rich in polyphenols and other micronutrients, and were identified as potential lines that could be utilized and promoted as additional source of food supply in sub-Saharan Africa.

Many chapters as well as projects presented in the Appendix have been published as this dissertation was completed [1-5] An additional few have been submitted for publication. Built upon the current momentum, more exciting work remains to be achieved.

In chapter IV, a high-throughput and sensitive method was established for amino acid analysis, and was used for determination of free amino acids in leafy AIVs. However, the current progress is not enough to understand the leafy protein nutrition. One of the important future work is to determine the proteinaceous amino acids profile so as to unveil the total protein content and the associated protein quality. It is of excitement to note that on one hand the total content of *free* amino acids in leafy AIVs is high from several to ten percent of dry mass, and on the other, the total crude protein in leafy AIVs have also been reported in literature (by Kjeldahl's method) to be satisfactorily high at many tens of percent (up to 70%) of the dry mass. An amino acid-based accurate determination of the total protein content and particularly the protein quality is likely to reveal high value in use of AIVs as an important source of protein. An online interactive dashboard using R Shiny, etc., could be conveniently set up to calculate dietary ratio of AIVs to make a complete protein formula. The protein quality data of household vegetables apart from AIVs could be readily accessed from the USDA database (*https://data.nal.usda.gov/dataset/usda-national-nutrient-database-standard-reference-legacy-release*), which should also be incorporated into this online dashboard. This online program could be readily applied in

sub-Saharan Africa to guide African people's day-to-day dietary practice for improved protein nutrition.

Based on this idea, one specific proposal is to formulate an AIV protein shake. With increasingly more people nowadays turn to a healthier and greener diet, vegetable smoothies are getting amazingly popular too. Commercially available cold-pressed bottled veggie-smoothies are available on the market sold at high price. AIV-based smoothies are not only green, but also high in protein, meanwhile also being an ideal supplement for vegetarians. This project could be one of prospective commercial potential.

Apart from a nutritional perspective, in Chapter IV machine learning (ML) methods were used for AIV classification purpose based on free amino acids profile. More work would be done for improvement of prediction accuracy. One way is to further fine-tune each model already in use, and also to test more ML methods. Another important approach is to ensemble different machine learning methods into a single method, drawing upon the respective strengths of each model at different sample space. Interpretation of ML methods is an emerging and rapidly developing branch and filed of ML, which turns the ML black box into a more understandable and transparent mechanism. It is of excitement to use these many new techniques to further enhance the work presented in this dissertation.

In chapter VI, the structure of potentially novel glycoalkaloids were tentatively identified using mass spectrometry as aforementioned. However, the exact structure remains to be further clarified and confirmed. Fractionation and purification of individual compounds by column chromatography and following NMR study is needed. Toxicological and medicinal activities of the glycoalkaloids, in form of individual compounds, as mixture, or in form of glycoalkaloid-enriched berry extract, etc., remains to

be investigated. The bioactivity comparison between the glycosidic form and the free aglycone is also interesting and important to study. The knowledge regarding these issues provides insight and guidance to consumption safety and potential medicinal application.

In Chapter VII and VIII, solamargine as the only commercially available standard was used to construct calibration for quantification of all glycoalkaloids in the berries. Such quantitation remained approximate as small changes in glycoalkaloidal structure may lead to noticeable difference in instrumentational signal response, and as such solamargine may not necessarily be representative for all glycoalkaloids. After collection of individual glycoalkaloids, more standards will be available to use for improved quantitation accuracy. Re-determination of the content of glycoalkaloids in the nightshade berries using the harvested standards may be beneficial. In addition, signal comparison between new glycoalkaloids vs. solamargine provides insight into the association of chemical structures and signal response, which could be applied for correction of quantification when solamargine is used as the sole standard (after all it's the only commercial standard conveniently available).

With the harvest of the many individual novel glycoalkaloids, another topic of interest is a re-visit of the fragmentation pathway of glycoalkaloids. While the basic fragmentation pattern of the glycoalkaloidal skeleton (here particularly regarding solamargine-like compounds) has been well investigated in both literature and this dissertation, many nuances remains unclear, such as the fine-tune of this pattern (regarding intensity and m/z) induced by the various substitution on the spirosolane skeleton and stereo-isomerization. Rationalization of the observed fragments of the many new glycoalkaloids under different collision energy using high-resolution mass spectrometry can produce more insight into

mass spectra interpretation, and provides a solid foundation for structural identification of new glycoalkaloids to be found in the future.

References:

[1] Yuan B, Byrnes D, Giurleo D, Villani T, Simon JE, Wu Q. Rapid screening of toxic glycoalkaloids and micronutrients in edible nightshades (*Solanum* spp.). J Food Drug Anal 2018;26:751-760.

[2] Yuan B, Byrnes D, Dinssa FF, Simon JE, Wu Q. Identification of Polyphenols, Glycoalkaloids, and Saponins in *Solanum scabrum* Berries Using HPLC-UV/Vis-MS. J Food Sci 2019;84:235-243.

[3] Yuan B, Byrnes D, Dinssa FF, Simon JE, Wu Q. Quantity assessment of polyphenols, glycoalkaloids and saponins in *Solanum scabrum* berries of different genetic sources and maturity by HPLC/UV-visible/MS methods. J Sci Food Agric 2019;99:3578-3587.

[4] Yuan B, Dinssa FF, Simon JE, Wu Q. Simultaneous quantification of polyphenols, glycoalkaloids and saponins in African nightshade leaves using ultra-high performance liquid chromatography tandem mass spectrometry with acid assisted hydrolysis and multivariate analysis. Food Chem 2020;312:126030.

[5] Yuan B, Zhao D, Du R, Kshatriya D, Bello NT, Simon JE, Wu Q. A highly sensitive ultra-high performance liquid chromatography/tandem mass spectrometry method with in-source fragmentation for rapid quantification of raspberry ketone. J Food Drug Anal 2019;27:778-785.

APPENDIX A. SUPPLEMENTARY TABLES & FIGURES

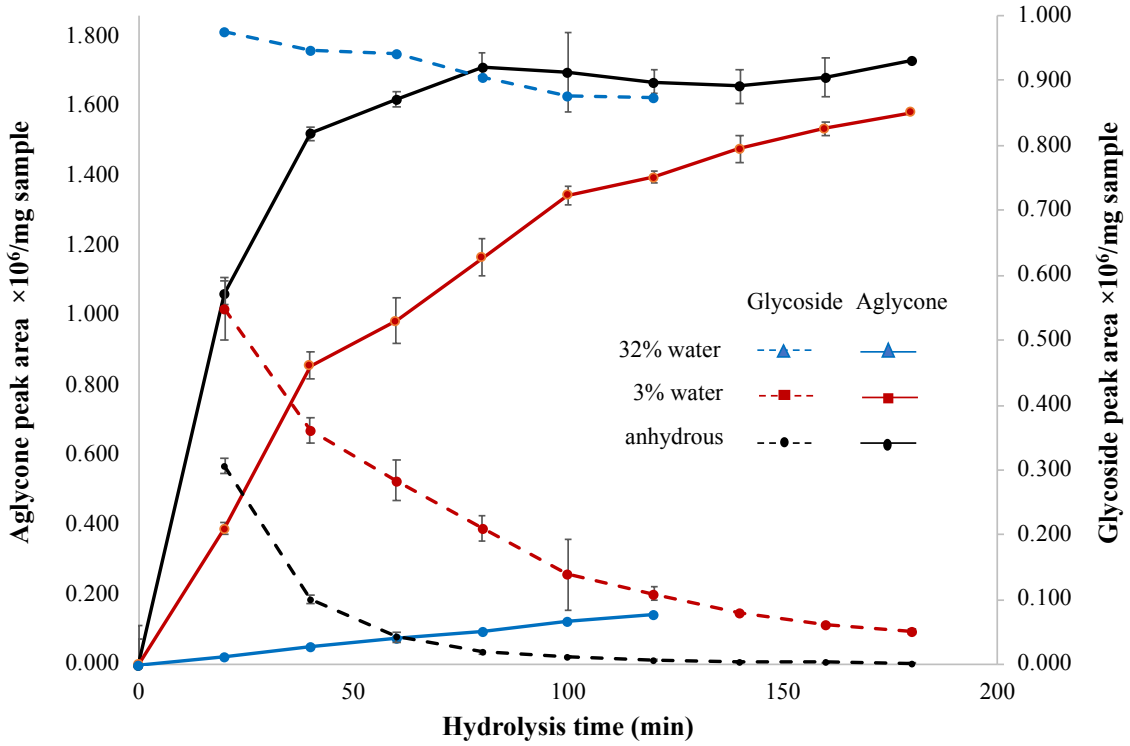


Figure A-1. Glycoalkaloid hydrolysis profile of *Solanum nigrum* fruit. Hydrolysis was conducted at 70°C using 0.5 M HCl methanol with different water percent in solvent. For solvent with 32% (vol. percent) water, 0.9 ml concentrated hydrochloric acid (CHA) (35~37%, mass percent, ~1.2g/mL) was mixed with 20 ml 70:30 (v/v) methanol/water. For solvent with 3% water (vol. percent), 0.9 ml CHA was mixed with 20 ml methanol. For anhydrous solvent, CHA was dripped into heated concentrated sulfuric acid (CSA), and hydrogen chloride generated was passed through a secondary flask filled with CSA for additional dehydration, and then channeled into chilled methanol to make a stock solution. The stock was then diluted with methanol to 0.5 M HCl determined by titration using standardized 0.1 M sodium hydroxide solution. Standard deviation was based on two replicates. Glycosides refer to solamargine, solasonine and corresponding di- or mono-glycosylated counterparts. The aglycone refers to solasodine.

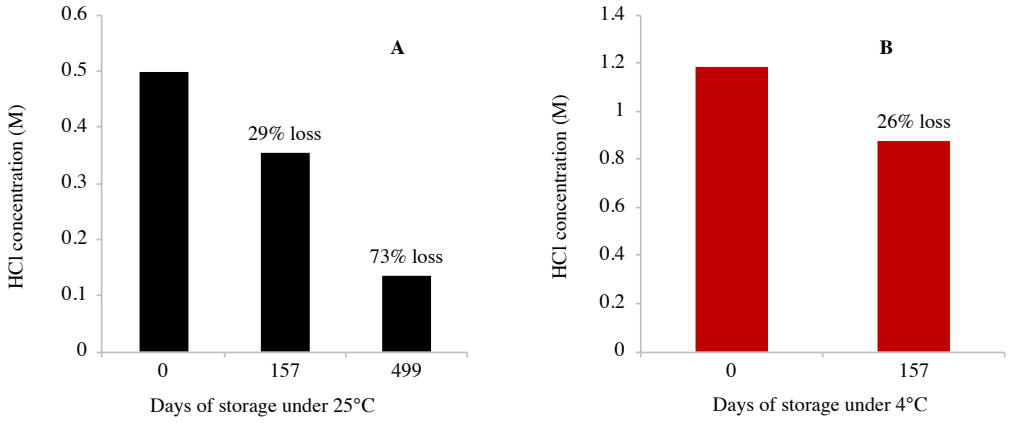


Figure A-2. Hydrochloric acid loss in methanol solution during storage in different conditions.

Table A-1. Moisture content in methanol solution of 1M $[H]^+$ prepared by blending methanol with concentrated hydrochloric acid (37%) or sulfuric acid (98%).

source of acid	acid molarity (M)	concentrated acid (g)	moisture (g)	moisture percent
hydrochloric acid	1.0	98.6	62.15	6.21%
sulfuric acid	0.5	50.0	1.00	0.10%

Calculation is based on 1L acidified methanol solution.

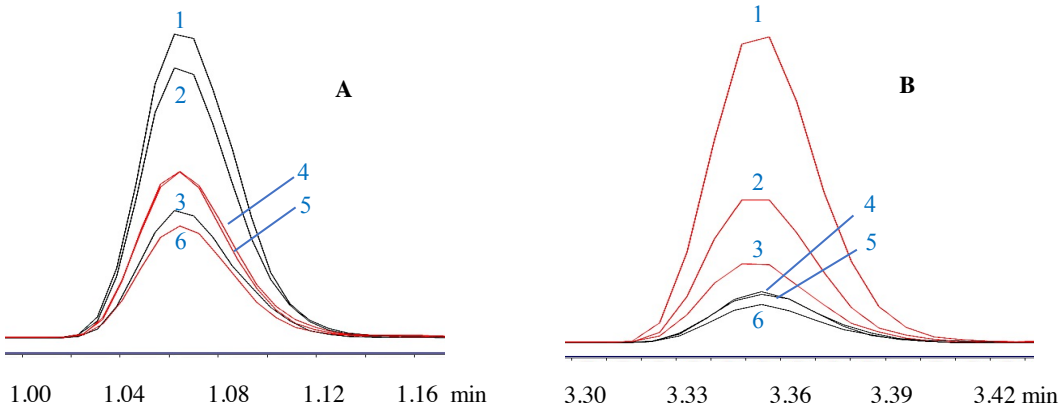


Figure A-3. In-source fragmentation of solamargine (**A**) and dioscin (**B**). Chromatograms (overlaid) were acquired using method (a) but in multiple reaction monitoring (MRM) mode, not in dynamic MRM (dMRM). Peaks with the precursor ion being the parent glycosides are noted in black, and peaks with the precursor ion being the in-source

produced aglycone ion noted in scarlet. Notice the lack of peak smoothness due to fewer data points acquired across peak elution time using MRM instead of dMRM. MRM transitions are separately listed below for easy reading.

Solamargine MRMs:

A1, 868.5 -> 129.0; **A2**, 868.5 -> 850.5; **A3**, 868.5 -> 253.5;

A4, 414.3 -> 157.0; **A5**, 414.3 -> 253.1; **A6**, 414.3 -> 396.4;

Dioscin MRMS:

B1, 415.3 -> 271.2; **B2**, 415.1 -> 253.1; **B3**, 415.1 -> 157.0;

B4, 869.5 -> 253.1; **B5**, 869.5 -> 146.9, **B6**. 869.5 -> 271.0.

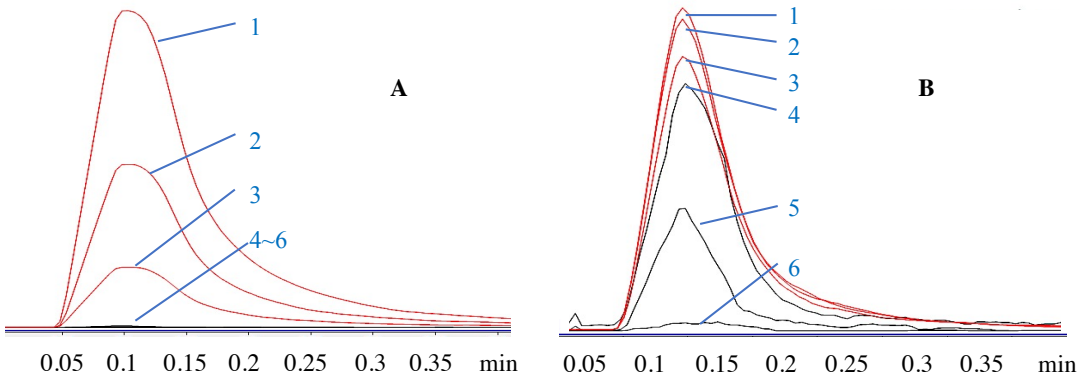


Figure A-4. Higher signal response of quercetin (**A**) and kaempferol (**B**) acquired under negative (red peak) than positive polarity (black peak). Union was used in replacement of column for rapid elution. Three most intense MRM chromatograms (overlaid) under each polarity are shown for each compound. MRM transitions are shown below; collision energy (CE) for positive MRMs are noted in parenthesis (in unit eV), and for the negative refers to **Table III-2**.

Quercetin MRMs,

A1, 301.0 -> 150.9; **A2**, 301.0 -> 178.9; **A3**, 301.0 -> 121.0;

A4, 303.1 -> 153.0 (36); **A5**, 303.1 ->229.0 (32); **A6**, 303.1 ->257 (28).

Kaempferol MRMs,

B1, 285.0 ->93.0; **B2**, 285.0 -> 159.0; **B3**, 285.0 ->186.9

B4, 287.0 -> 120.9 (36); **B5**, 287.0 -> 153.0 (40); **B6**, 287.0 -> 230.8 (28).

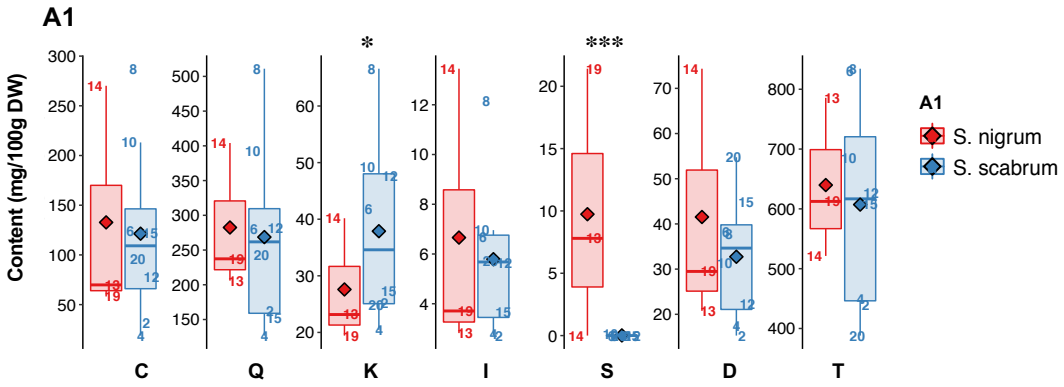


Figure A-5. Contrast analysis of *S. nigrum* vs. S. *scabrum* species mean content. Sample 16 was excluded from calculation compared with **Figure III-4**, resulting in insignificant difference in the mean content of diosgenin ($p > 0.05$) and tigogenin ($p > 0.5$) of the two species. Sample 16 may be considered as an "outlier" due to its peculiar profile, with its unique rhamnetin and the unusual highest content in diosgenin and lowest tigogenin among all samples investigated. *, $p < 0.05$; ***, $p < 0.001$. Compounds are abbreviated into initial letters; refer to **Table III-2**.

Table A-2. Condition and formula for preparation of the mobile phase, testing the effect of different ammonium formate concentration in the mobile phase upon chromatographic performance.

Mobile phase	AF conc. (mM)	0.01	0.05	0.1	0.25	0.5	1	5
A (water phase), with 0.15% FA	FA vol. (mL)	0.6	0.6	0.6	0.6	0.6	0.6	0.6
	AMF stock vol. (µL)	21	105	210	525	1050	2100	10500
	After mixing FA and AMF stock solution, use water to bring the final volume to 400 mL.							
B (organic phase), with 0.15% FA, *ca.* 95% acetonitrile.	FA vol. (mL)	0.6	0.6	0.6	0.6	0.6	0.6	0.6
	AMF stock vol. (µL)	21	105	210	525	1050	2100	10500
	water vol. (mL)	16.8	16.7	16.6	16.275	15.75	14.7	6.3
	After mixing FA, AMF stock solution and water, use acetonitrile to bring the final volume to 400 mL.							

For abbreviations, AMF, ammonium formate; FA, formic acid. Note that in this part of method development, both mobile phase A and B were buffered with AMF. Addition of AMF to only the water phase not in the organic phase, though commonly seen in literature, was not considered efficient in this work as the buffer imbalance could cause difficulty in column equilibration in gradient elution and thus requires long time of column equilibration.

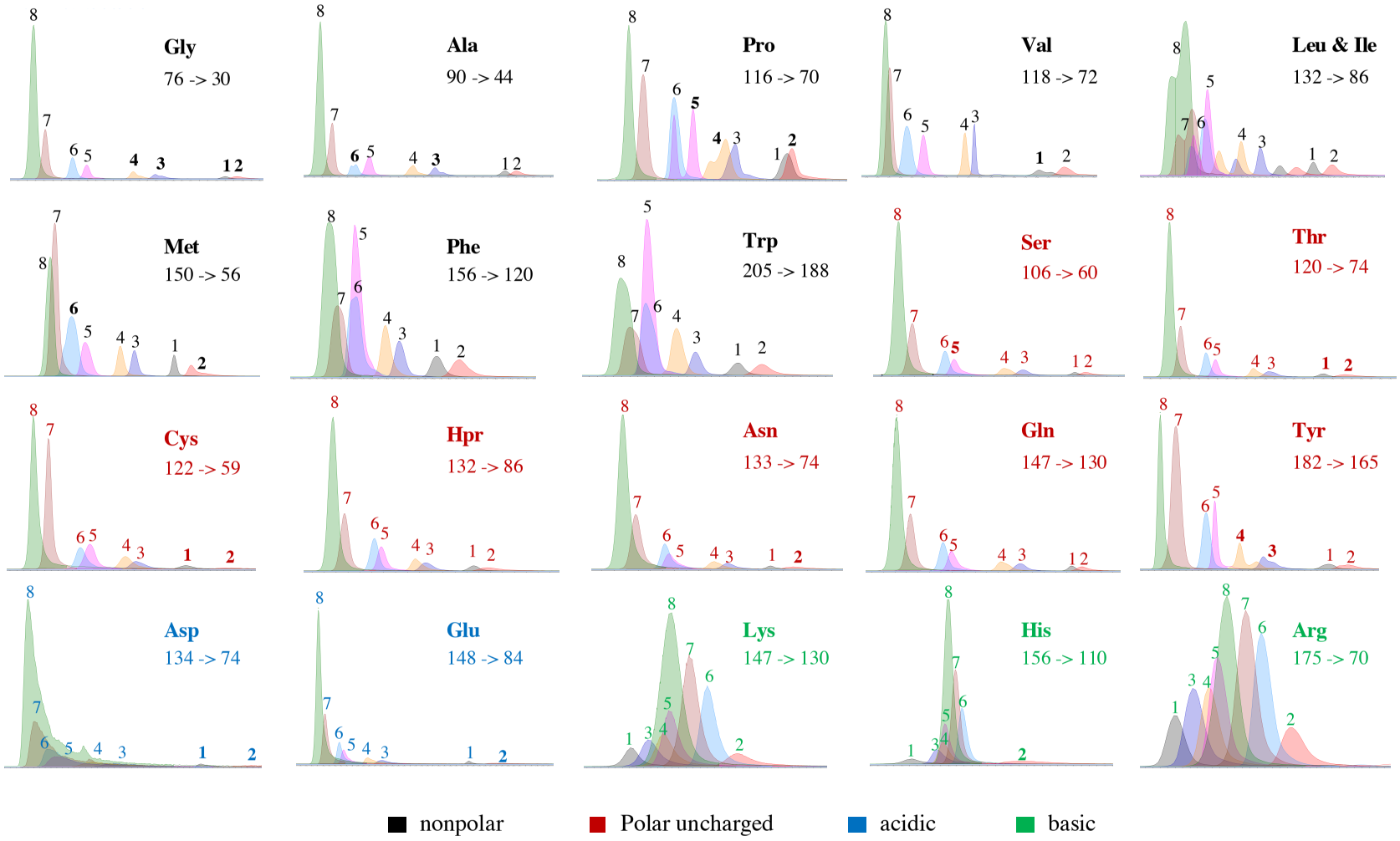

Gly
76 -> 30
Ala
90 -> 44
Pro
116 -> 70
Val
118 -> 72
Leu & Ile
132 -> 86
Met
150 -> 56
Phe
156 -> 120
Trp
205 -> 188
Ser
106 -> 60
Thr
120 -> 74
Cys
122 -> 59
Hpr
132 -> 86
Asn
133 -> 74
Gln
147 -> 130
Tyr
182 -> 165
Asp
134 -> 74
Glu
148 -> 84
Lys
147 -> 130
His
156 -> 110
Arg
175 -> 70
nonpolar
Polar uncharged
acidic
basic

Figure A-6. Overlaid multiple reaction monitoring (MRM) chromatograms of amino acids using different concentrations of ammonium formate in the mobile phase. Product ions of the most abundance (note: not necessarily the quantifier ion. See manuscript *section 3.2* for more discussion) were shown. The binary mobile phase was composed with water and acetonitrile (ACN), both with 0.15 % formic acid (the finalized method used 0.1% formic acid), with more specifications referring to **Table A-2**. Amino acid standard mixture was used and prepared in 50/50 water/ACN (finalized method used 10/90 water/ACN with 10 mM HCl) with 3 µL injection. Column was thermostatted at 30 °C. Buffer concentration was coded with: 1, 5mM; 2, 1mM; 3, 0.5 mM; 4, 0.25 mM; 5, 0.1 mM; 6, 0.05 mM; 7, 0.01 mM; 8, buffer free. The buffer notes are bolded for peaks remarkably distorted or with wide width. Note the remarkable bifurcation of proline at 0.1 mM buffer condition (peak 5), as well as similar phenomena of tyrosine at 0.25 and 0.5 mM (peak 3 and 4) and alanine at 0.05 mM buffer concentration (peak 6). Also note that for isobaric compounds leucine and isoleucine, the secondary qualifier ion transition 132->86 was used for demonstration of coelution and decreasing chromatographic resolution with lower buffer concentration. The two compounds were distinguished from one another by using different quantifier product ions unique to each other in the final optimized method.

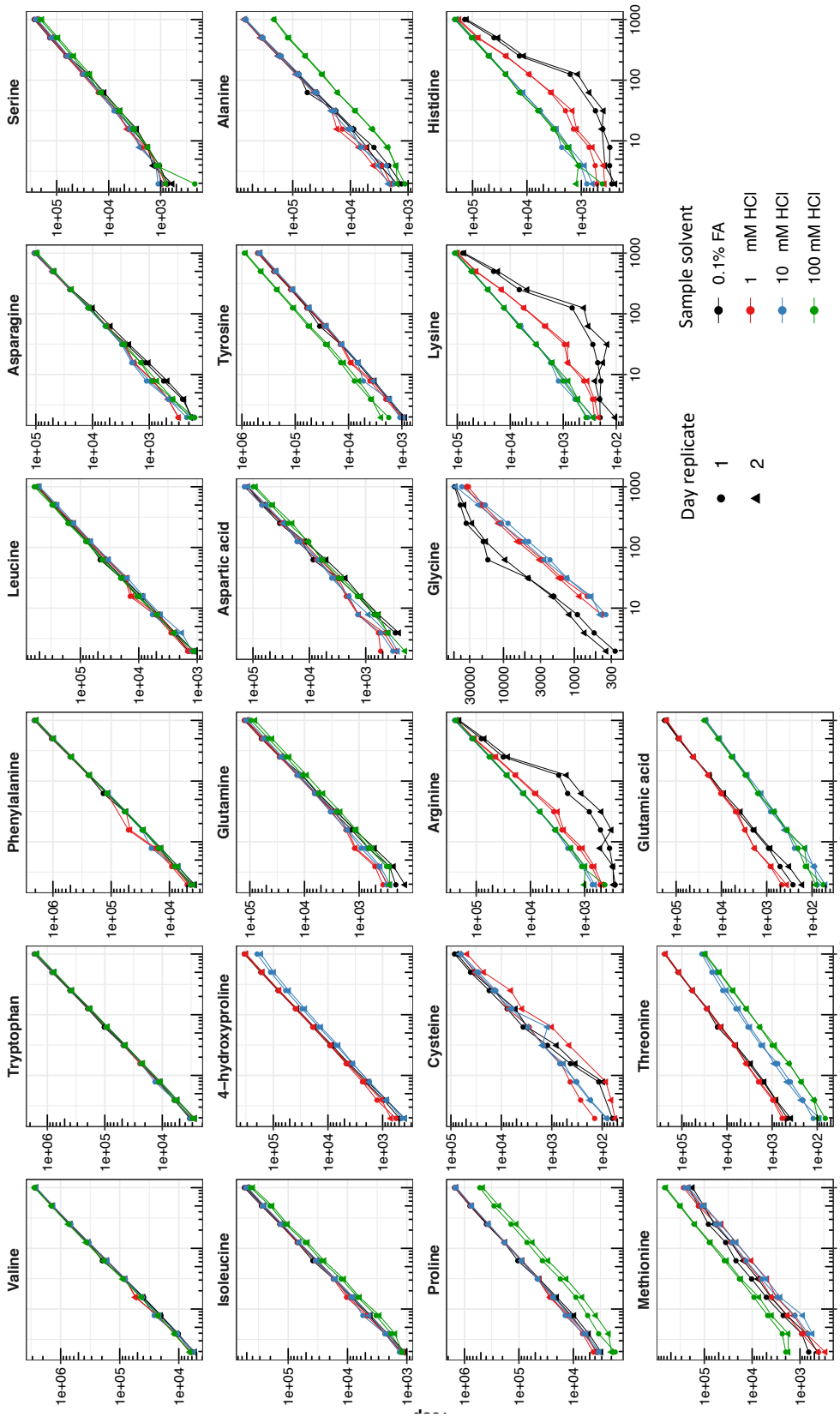

Peak area
Nominal concentration (µg/mL)
Day replicate
1
2
Sample solvent
0.1% FA
1 mM HCl
10 mM HCl
100 mM HCl
Valine
Tryptophan
Phenylalanine
Leucine
Asparagine
Serine
Isoleucine
4−hydroxyproline
Glutamine
Aspartic acid
Tyrosine
Alanine
Proline
Cysteine
Arginine
Glycine
Lysine
Histidine
Methionine
Threonine
Glutamic acid

Figure A-7. Sample solvent effect on amino acids' calibration. Compounds are arranged in order of increasing susceptibility to sample solvent composition. The sample solvent used was 90 % acetonitrile (ACN) with different acid composition. 90 % ACN with 0.1% formic acid (FA) was the starting mobile phase composition. Injection of each calibration set was replicated over two days, during which time the samples were stored in 4 °C. For some compounds such as glycine, cysteine and 4-hydroxyproline, etc., the results under 100 mM HCl are not shown as the retention time shifted (due to solvent effect, see **Figure IV-2**) outside dMRM time zone in the experiment and thus not detected.

Table A-3. Validated accuracy of the HILIC-MS/MS method for amino acid analysis. Accuracy was validated at seven spike levels noted as A to G. The values are presented in unit of percentage (%).

No.	compounds	A	B	C	D	E	F	G
1	4-hydroxyproline	107.4 ± 2	106.9 ± 1.1	107.3 ± 1.9	107.5 ± 1.6	106.1 ± 1	105 ± 4.9	111.4 ± 9.8
2	alanine	108.1 ± 3.1	111.7 ± 1.5	113.4 ± 3.2	114.6 ± 2.3	115.8 ± 5.9	110.7 ± 8.5	108.3 ± 22.9
3	arginine	109.9 ± 3.5	106.2 ± 1.7	108.4 ± 2.7	109.4 ± 2.2	111.6 ± 4.1	114.3 ± 13	108.4 ± 17.5
4	asparagine	110.1 ± 2.7	108.8 ± 1.4	111 ± 1.9	112.6 ± 2.3	104.9 ± 9.4	117.8 ± 14.2	104.3 ± 17.4
5	aspartic acid	116.9 ± 1.1	115.6 ± 0.7	114.9 ± 2.7	110.7 ± 3	111.6 ± 5.5	108.4 ± 18.4	98.7 ± 25.3
6	cysteine	106.8 ± 3	110.9 ± 1.8	108.5 ± 1.9	109.8 ± 1.9	104.6 ± 3.8	103.8 ± 3	106.5 ± 4.7
7	glutamic acid	102.8 ± 4	95.3 ± 5.3	93.3 ± 9.5	96.7 ± 3.8	92.6 ± 5.7	93.1 ± 21.8	75 ± 32.5
8	glutamine	100.4 ± 2.7	101.4 ± 0.9	101.3 ± 2.6	106.1 ± 3	104.4 ± 3.4	100.7 ± 5.3	110.5 ± 15.6
9	glycine	109.1 ± 2.9	122.3 ± 1.1	124.4 ± 3.7	123.3 ± 4.2	120.9 ± 5.9	121.5 ± 24.1	102.6 ± 17.4
10	histidine	111.8 ± 3	102.9 ± 2.4	109.1 ± 2.8	111 ± 3.2	105.1 ± 10.2	94.3 ± 17.2	121.2 ± 32.3
11	isoleucine	105.4 ± 6.2	110.2 ± 1.4	112.4 ± 2.6	114.1 ± 2.8	117.2 ± 4.6	115.9 ± 5.7	122.8 ± 14.8
12	leucine	107.2 ± 1.9	108.1 ± 1.4	109 ± 1.9	109.9 ± 3	110.4 ± 3.4	105.4 ± 12.8	109.3 ± 16.6
13	lysine	110.1 ± 2.8	105.6 ± 1.9	108 ± 3.9	107.6 ± 3	106.8 ± 4.9	101.7 ± 18.5	82.9 ± 16.6
14	methionine	107.4 ± 4	109.5 ± 3.2	106.8 ± 3.7	106.3 ± 3.5	107.6 ± 3.3	103 ± 4.1	105.8 ± 7
15	phenylalanine	107.3 ± 2.6	107.2 ± 0.2	107.9 ± 4	108.5 ± 1.8	107.4 ± 1.8	105.7 ± 3.3	109.3 ± 14
16	proline	107.4 ± 2	110.6 ± 0.9	111.6 ± 4.5	113.1 ± 4.8	111.7 ± 8.6	112 ± 19.2	103.3 ± 52
17	serine	104.9 ± 2.7	101.1 ± 2.2	98.6 ± 2.3	92.5 ± 3.6	79.9 ± 3.7	82.7 ± 6.7	74.8 ± 15.2
18	threonine	101.4 ± 4.2	93.9 ± 1.5	91 ± 2.6	93.6 ± 0.9	91 ± 1.3	84.9 ± 4.2	96.6 ± 5
19	tryptophan	107.7 ± 3.3	107 ± 0.9	106.1 ± 3.3	106.9 ± 1.7	105.2 ± 0.8	105.5 ± 1.5	105.6 ± 7.1
20	tyrosine	107.8 ± 2.4	108 ± 1.1	107.5 ± 2.3	107.1 ± 0.9	106.5 ± 2.1	106.2 ± 4	106 ± 7.5
21	valine	108.6 ± 2.5	109.5 ± 2.5	110.6 ± 3.3	112.2 ± 1.9	110.2 ± 5.7	107.4 ± 6.8	111.8 ± 18

Table A-4. Validated matrix effect of the HILIC-MS/MS method for amino acid analysis. Matrix effect was validated at six spike levels noted as A to E and G. The values are presented in unit of percentage (%).

No.	compounds	A	B	C	D	E	G
1	4-hydroxyproline	101.9 ± 3.3	102.5 ± 5.3	101.1 ± 3.5	96.1 ± 6.9	103.7 ± 3.3	112 ± 16.3
2	alanine	99.7 ± 3.9	103.8 ± 4.6	100.9 ± 4.6	97.4 ± 6.1	106.4 ± 6	102.7 ± 22.4
3	arginine	100.4 ± 4.3	99.6 ± 4.7	101.9 ± 3.4	97.4 ± 7.9	104.6 ± 7.5	99.8 ± 18.9
4	asparagine	100.2 ± 4.3	102.7 ± 4.6	100.4 ± 4.6	98 ± 7	97.8 ± 12.3	116 ± 24.9
5	aspartic acid	99.7 ± 2.3	100.7 ± 4.2	102.7 ± 4.9	96.9 ± 9.8	98.5 ± 5	74.3 ± 19.9
6	cysteine	99.3 ± 4.2	105.4 ± 4.5	100.8 ± 3.3	98 ± 5.8	97.6 ± 4.7	107.3 ± 13.9
7	glutamic acid	106.4 ± 7.3	109.2 ± 6.4	100.5 ± 11.3	115.2 ± 22.8	105.7 ± 12.2	83.8 ± 37.8
8	glutamine	100.9 ± 4.2	102.9 ± 4.8	101.4 ± 4.2	103.3 ± 8.2	103.6 ± 5.8	124.8 ± 23.5
9	glycine	98.3 ± 3.9	111.6 ± 4.6	106.8 ± 5.4	100.4 ± 7.6	109.9 ± 9.7	110.1 ± 21.2
10	histidine	102.2 ± 4.6	94.7 ± 6.1	102.1 ± 5.2	100.8 ± 10.3	96.6 ± 12.7	82.9 ± 24.6
11	isoleucine	97.2 ± 6.1	101.3 ± 5.1	101.1 ± 7.8	96.4 ± 7	100.5 ± 4.8	121 ± 14.7
12	leucine	96.5 ± 2.6	96 ± 4.1	91.3 ± 3.8	86.1 ± 6.9	91 ± 3.7	86.3 ± 15
13	lysine	100.6 ± 4	98.8 ± 4.1	100.3 ± 3.8	92.5 ± 7.5	92.8 ± 6.5	57.5 ± 12.5
14	methionine	103.8 ± 6.1	106.8 ± 7	99.6 ± 6.1	93.8 ± 8.7	102.5 ± 7.7	103 ± 8.9
15	phenylalanine	99.6 ± 3.9	100.8 ± 4.4	100.9 ± 5.2	97.8 ± 7.1	101.2 ± 2.8	107 ± 14.6
16	proline	98.8 ± 3.5	103.4 ± 4.5	101.1 ± 5.6	97.3 ± 7.3	102.9 ± 8.3	99.9 ± 50.3
17	serine	101 ± 3.4	101.7 ± 6.6	98.5 ± 4.3	94.9 ± 7.9	90.8 ± 5.6	61.3 ± 13
18	threonine	104.7 ± 5.2	101.4 ± 3.5	98.3 ± 4.5	97.4 ± 6.6	102.6 ± 3.1	134.9 ± 7.8
19	tryptophan	100.7 ± 3.7	101.7 ± 4.2	101.3 ± 4.4	97.9 ± 7.6	100.6 ± 2.3	105.4 ± 7.2
20	tyrosine	100.8 ± 2.9	103.2 ± 4.9	100.8 ± 4.1	97.1 ± 6.6	100.7 ± 2.7	105.9 ± 8.9
21	valine	102.2 ± 4.1	102.4 ± 4.7	100.6 ± 5.2	96.4 ± 6.9	101.1 ± 5.8	106.3 ± 17.3

Table A-5. Validated precision of the HILIC-MS/MS method for amino acid analysis. Precision was validated at seven spike levels noted as A to G. The values are presented in unit of percentage (%).

No.	compounds	A	B	C	D	E	F	G
1	4-hydroxyproline	1.13	1.78	0.95	1.92	4.45	5.33	7.35
2	alanine	1.45	1.93	1.05	2.66	1.66	2.54	1.52
3	arginine	0.93	0.69	0.79	2.20	4.63	5.95	1.43
4	asparagine	1.75	2.19	0.85	4.14	3.88	15.31	10.36
5	aspartic acid	1.68	2.29	2.25	7.58	5.99	9.89	3.59
6	cysteine	0.87	1.69	1.40	1.11	2.47	7.20	6.09
7	glutamic acid	1.35	2.80	1.80	4.38	3.28	14.81	6.51
8	glutamine	1.45	1.64	0.95	1.21	3.82	7.98	12.01
9	glycine	0.69	1.44	1.51	3.63	7.38	9.33	6.93
10	histidine	1.19	0.88	1.18	3.57	11.42	16.12	9.39
11	isoleucine	2.34	1.70	2.40	0.90	2.00	2.76	5.34
12	leucine	1.95	2.40	2.89	1.39	3.86	5.87	9.27
13	lysine	1.77	1.44	0.86	3.26	7.12	15.33	18.34
14	methionine	2.54	2.75	0.61	1.23	2.26	1.35	5.36
15	phenylalanine	1.58	1.26	1.38	0.76	1.54	3.42	3.16
16	proline	2.07	2.22	0.70	1.42	2.34	1.59	2.56
17	serine	2.95	1.87	1.55	7.75	5.13	11.24	3.41
18	threonine	0.22	2.25	2.03	1.51	5.60	4.52	9.00
19	tryptophan	0.80	2.26	2.05	1.37	2.68	0.46	2.07
20	tyrosine	1.26	2.08	0.64	1.54	1.89	2.93	3.04
21	valine	0.77	2.06	1.06	1.23	1.87	1.36	1.67

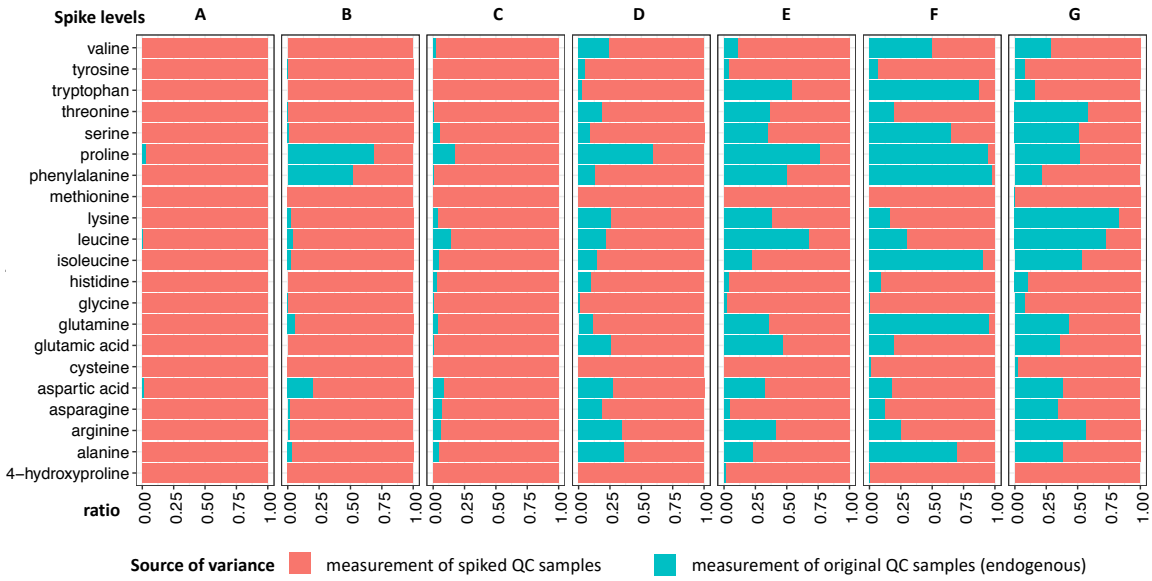


Figure A-8. Partition of the variance of accuracy at different spike levels. The error source is split to the measurement of the spiked quality control (QC) samples and that of the original unspiked QC samples.

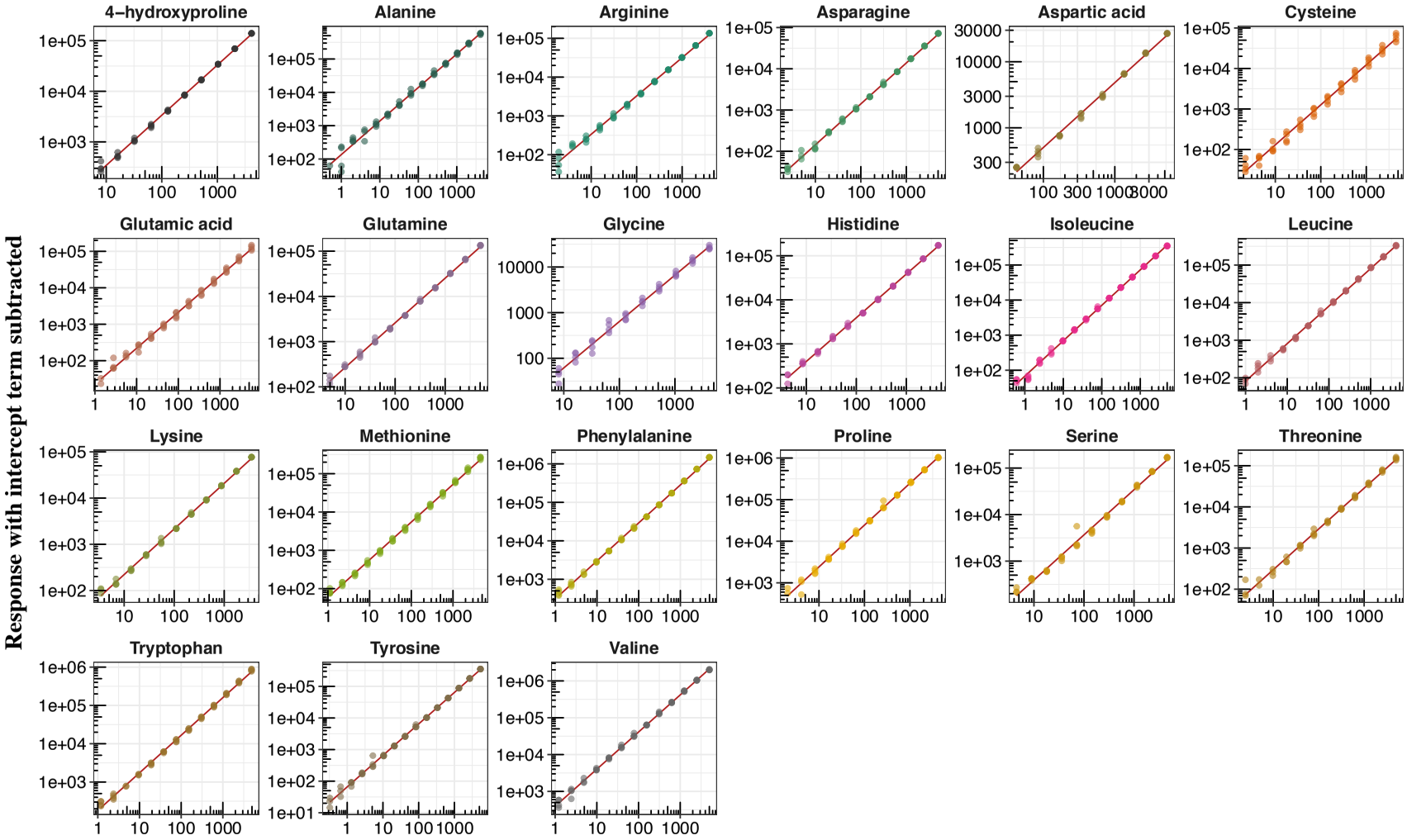

4−hydroxyproline
Alanine
Arginine
Asparagine
Aspartic acid
Cysteine
Glutamic acid
Glutamine
Glycine
Histidine
Isoleucine
Leucine
Lysine
Methionine
Phenylalanine
Proline
Serine
Threonine
Tryptophan
Tyrosine
Valine
Response with intercept term subtracted
Concentration (ng/mL)

Figure A-9. Visualization of calibration linearity range. To facilitate visualization across a liner range of three orders of magnitude, the intercept, computed based on regression with 1/x weight, is subtracted from peak response, and then both x and y scales are logarithmically transformed. That is, it is $\log(y - b)$ *(now the new y axis)* $= \log(ax) = \log(a)$ (*the new y-intercept*) $+ \log(x)$(*the concentration, with slope coefficient being one*) that is being plotted, instead of $y = ax + b$, with y being the response, a the slope coefficient, x the concentration, and b the y-intercept, both a and b regarding the original scale. Note that each of the four replicates of the calibration sets were prepared separately by serial dilution from the same stock solution.

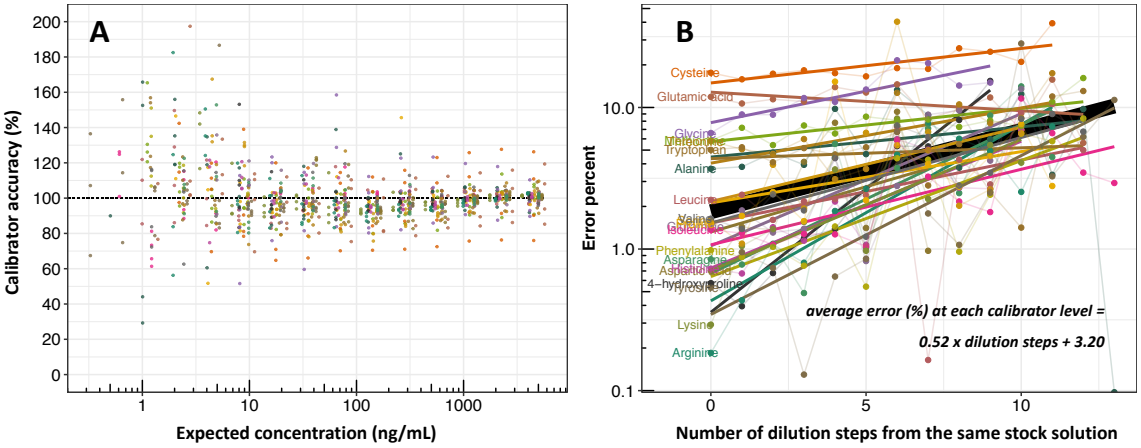


Figure A-10. Calibration accuracy and error analysis. Plot (**A**) shows the calibrator accuracy (CA) at each concentration level, with CA defined as the ratio of back-calculated concentration using constructed calibration curve vs. expected concentration. Four calibration sets were each prepared separately by serial dilution from the same stock solution. The CA plot reflects calibrators' linearity and consistency at each concentration level of the calibration curve. Note that the CA defined in this context is very different and clearly distinguished from the validated accuracy using spiked quality control samples as aforementioned. (**B**) shows the error propagation profile, with the error at each dilution step defined as the standard deviation of the four calibrator replicates' peak area divided by the average level. Each amino acid is separately regressed, and all points are jointly regressed represented by the central bold black regression line. The overall regression model follows a slope coefficient of 0.52, suggesting that with each additional step of dilution, an extra 0.52% of error percentage could be induced at that given level due to error propagation effect. This number gives a rough estimation of the researcher's consistency at using the pipette for serial dilution. In addition, this number also reflects the increasing susceptibility of integration consistency to decreasing concentration.

Actual	Predicted			
	Amaranth	Mustard	Nightshade	Spider plant
Amaranth	143	0	0	0
Mustard	***1***	86	0	***3***
Nightshade	***11***	0	128	0
Spider plant	***5***	***1***	0	166

Figure A-11. Confusion matrix of the linear discriminant analysis performed on ***all*** African indigenous vegetable samples. This confusion matrix is a summary of **Figure IV-7B**. *Note that in this analysis, there was no training-testing set split. The model was built up and tested on the entire AIV dataset.*

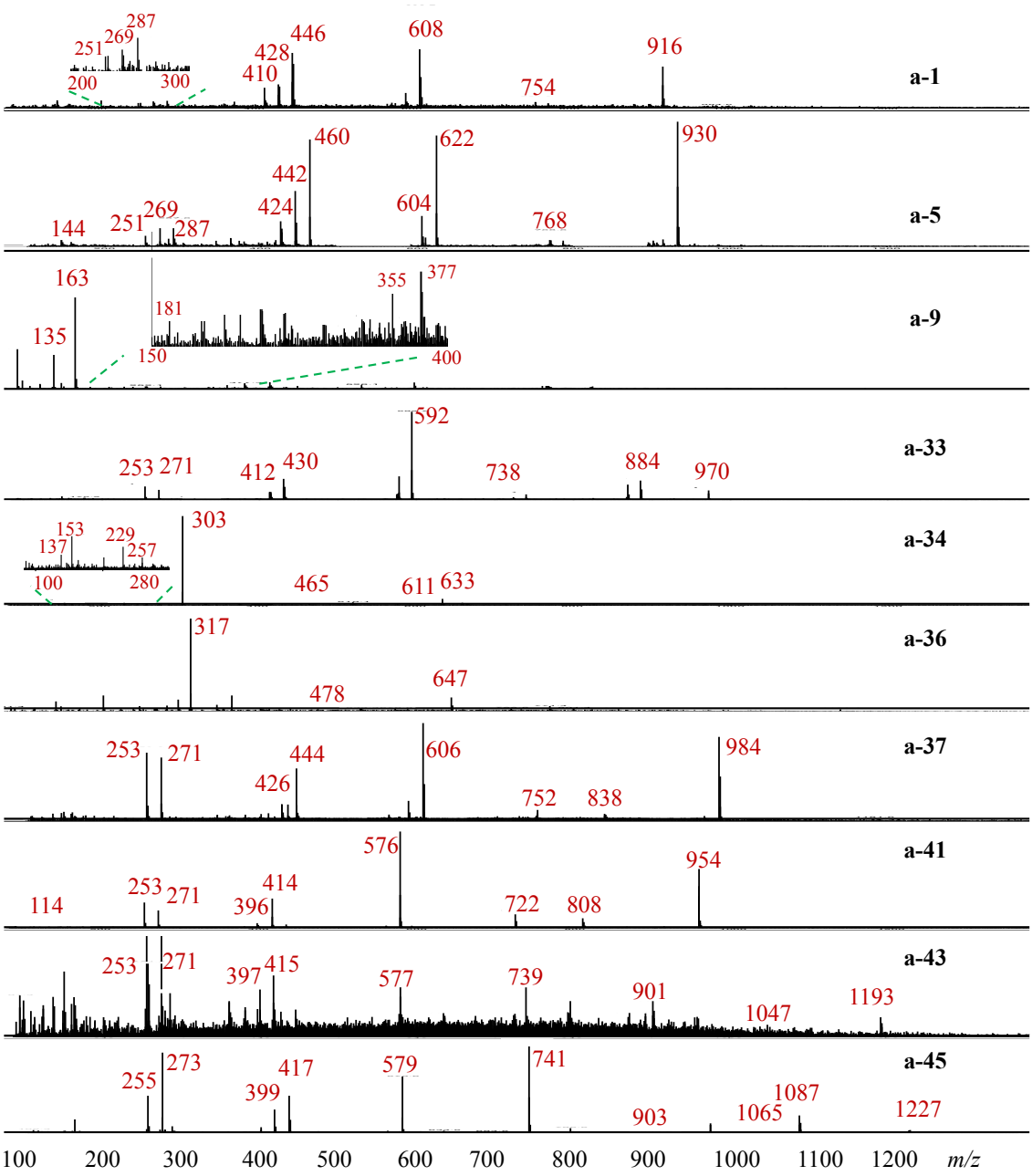


Figure A-12. Representative mass spectra of glycosides of different aglycones in the fruits of edible nightshade, *Solanum scabrum*, acquired by method (a). Each mass spectrum is labeled by the corresponding compound number in the upper right corner.

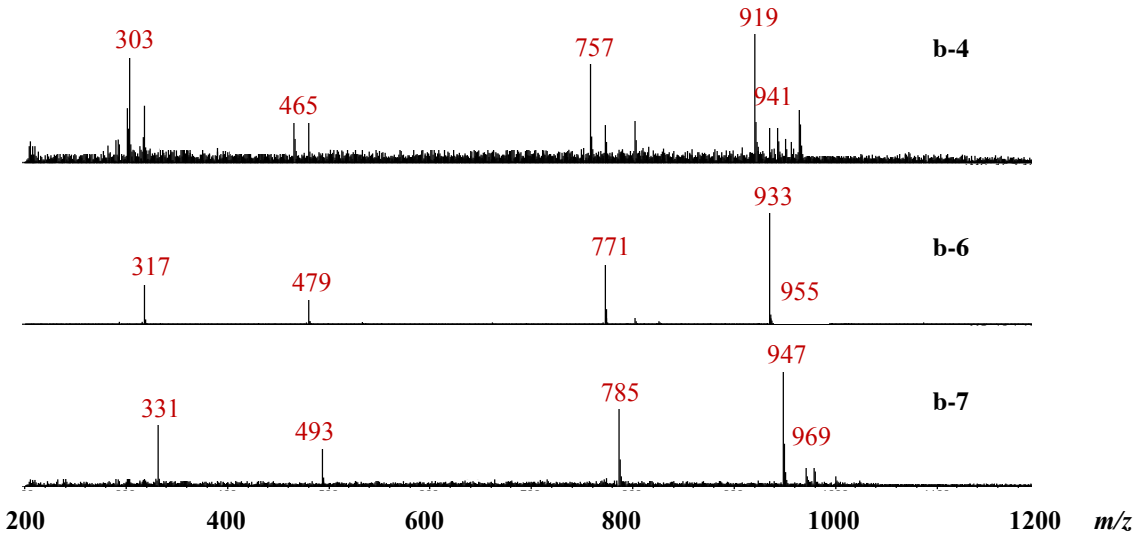


Figure A-13. Representative mass spectra of anthocyanins acquired by method (b). Each individual mass spectrum is labeled by the corresponding compound number in the upper right corner.

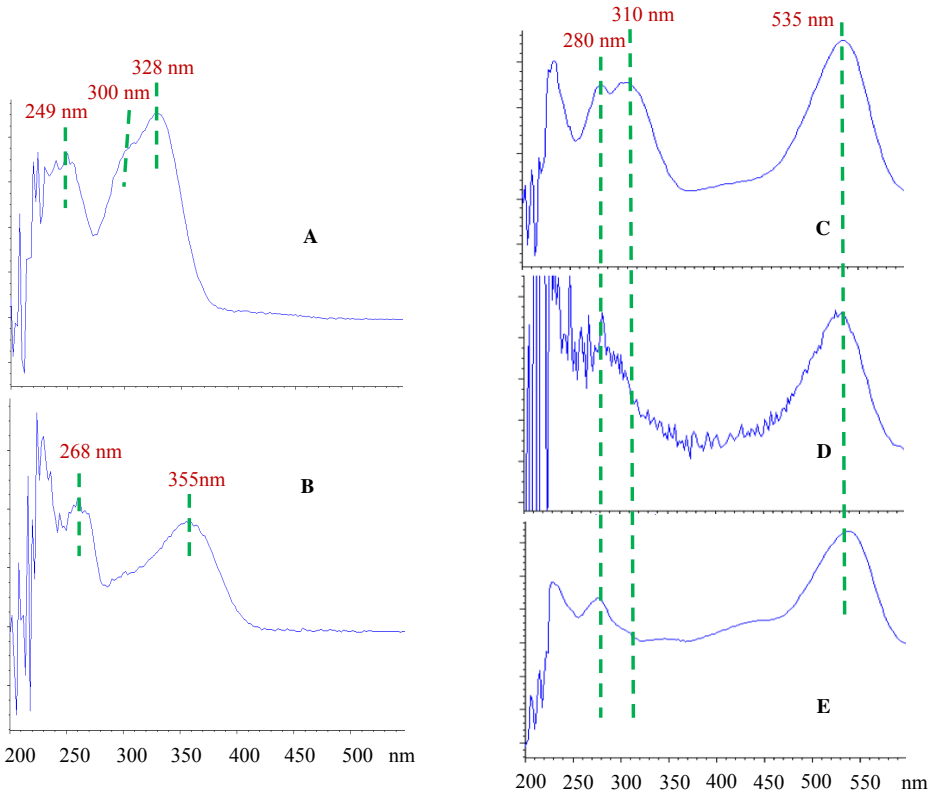


Figure A-14. UV-vis profile of neochlorogenic and chlorogenic acid (**a-6** and **a-9**) (**A**), quercetin glycoside (**a-34**) (**B**), petunidin-containing acylated anthocyanin (**b-3**) (**C**) and non-acylated simple anthocyanin (**b-1**) (**D**), and petunidin standard (**E**).

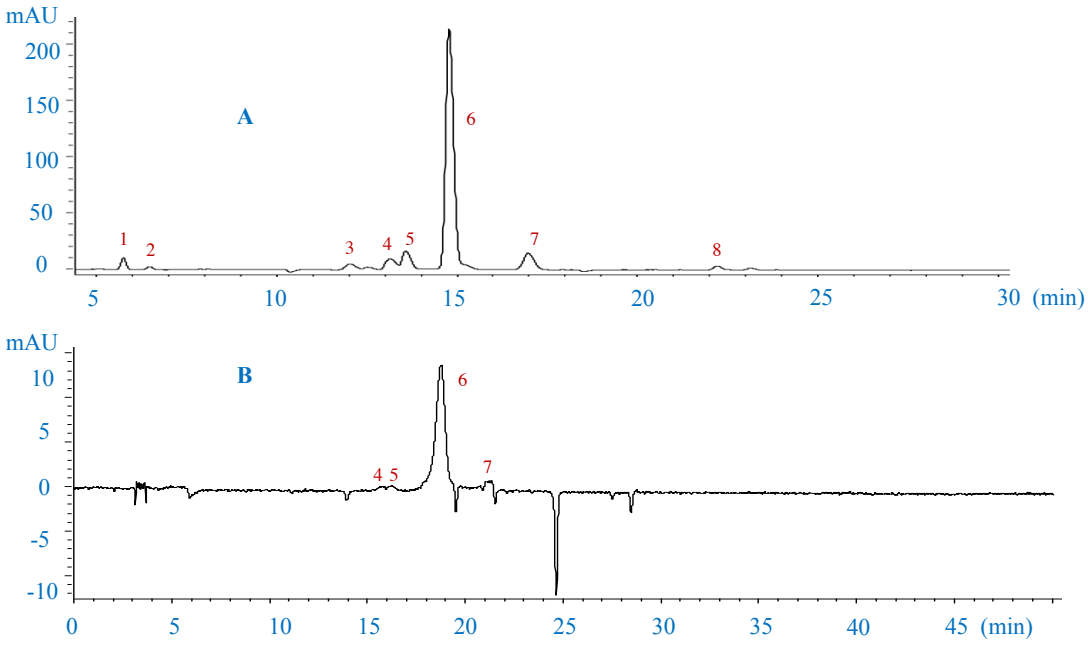


Figure A-15. UV chromatograms of anthocyanins in edible nightshade, *Solanum scabrum* fruits USDA PI 643126 under 520 nm using different mobile phase modifiers, i.e., (**A**) 0.4 % trifluoroacetic acid; (**B**) 0.1 % formic acid (FA). $W_{0.5}$, peak width is the full width at half maximum. A_s, asymmetry factor is distance from centerline to the back edge divided by the distance from the centerline to the front edge, measured at 10% maximum. N, theoretical plates numbers per column, is calculated as $N = 5.54 (t_R / w_{0.5})^2$, t_R being retention time.

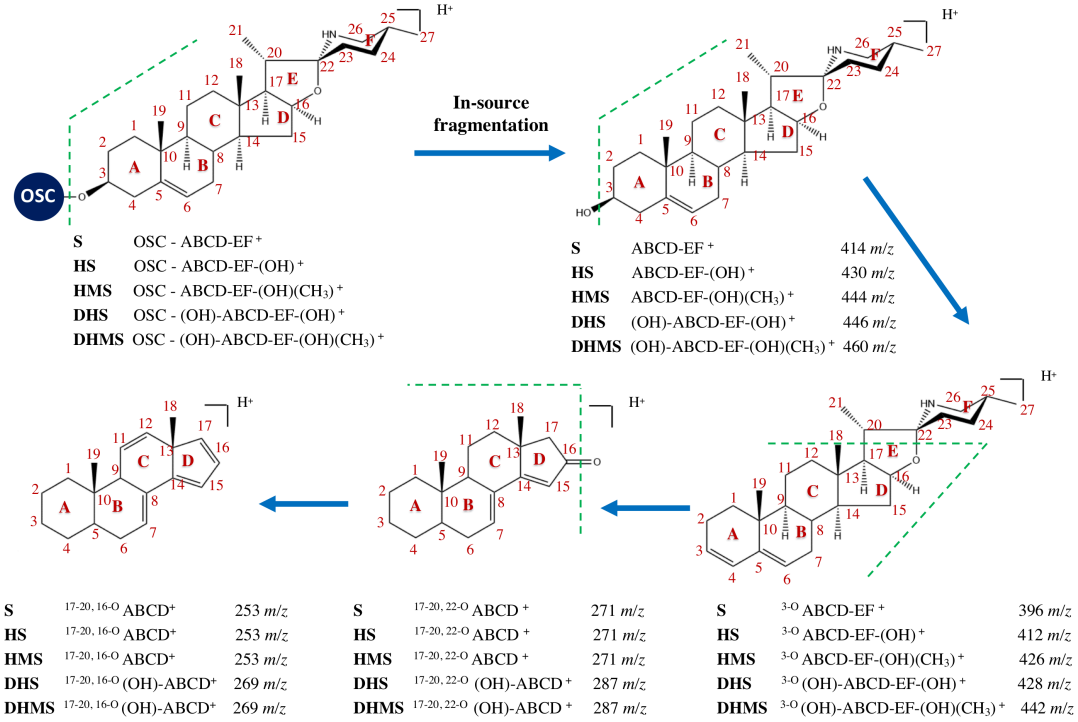


Figure A-16. Proposed fragmentation pathway and associated fragmental ions of glycoalkaloids. This figure is a more detailed specification based on prior **Figure VI-6**.

Table A-6. The central composite design (CCD) of drying gas temperature and sheath gas temperature.

standard order	design points	A (Coded)	B (Coded)	A (Actual)	B (Actual)	run order
1	factorial point	-1	-1	200	200	12
2		1	-1	300	200	2
3		-1	1	200	300	7
4		1	1	300	300	4
5	star/axial point	-1.414	0	179.3	250	10
6		1.414	0	320.7	250	6
7		0	-1.414	250	179.3	8
8		0	1.414	250	320.7	9
9	center point	0	0	250	250	11
10		0	0	250	250	5
11		0	0	250	250	1
12		0	0	250	250	3
13		0	0	250	250	13

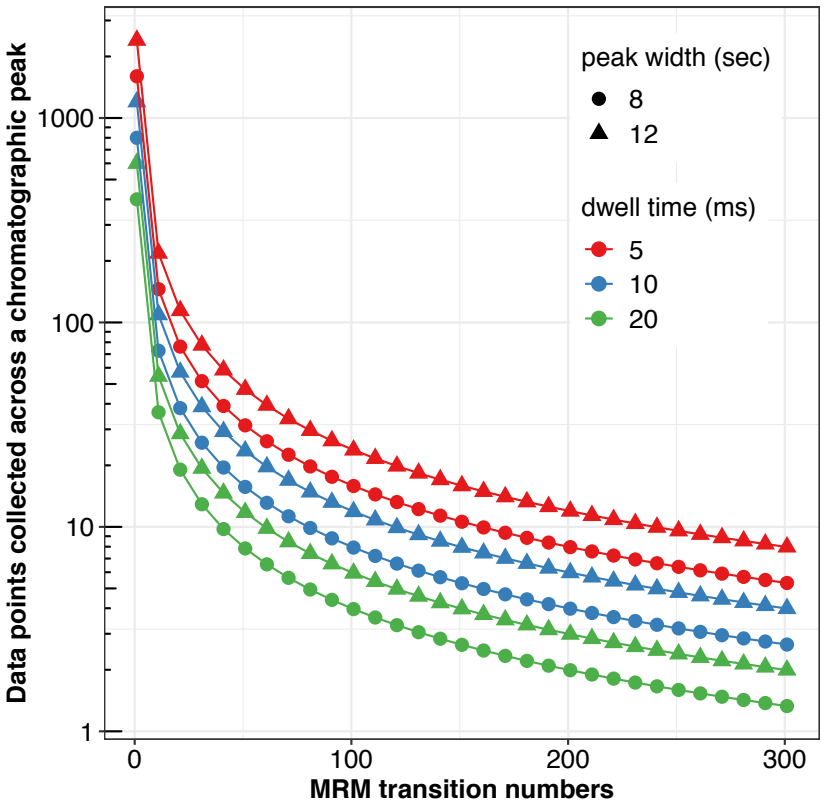


Figure A-17. Number of collectable data points across a chromatographic peak vs. number of MRM transitions under different dwell time and baseline peak width (BLPW). Regardless of its more rigid definition [5], the BLPW in this context for practical convenience simply refers to the baseline time range across the automatically / manually integrated chromatographic peak.

As demonstration of calculation, suppose that an MRM table contains a total of 100 compounds, each compound having 2 transitions, with each transitions dwell time 10 ms. As such, the total cycle or duty time to scan across all MRM transitions once would be approximately 100 (compounds) × 2 (transitions/compound) × 10 (ms/transition) = 2000 (ms), or 2 sec. Therefore, the number of data points that could be collected across a chromatographic peak of BLPW of 10 sec would be 10 sec / 2 sec = 5 data points, leading to very jagged unsmooth peak shape, and likely compromised sensitivity as well.

As a rule of thumb, the number of data points across a chromatographic peak should be no less than 15 to ensure good peak smoothness and symmetry. Reduction of dwell time to decrease the cycle time is a quick and convenient option to increase collectable data points. This adjustment nonetheless should be practiced with careful design. Based on our experience and current instrumental condition, the dwell time at best should be at least 10 ms; dwell time below 5 ms would compromise instrumental sensitivity noticeably.

APPENDIX B. RASPBERRY KETONE ANALYSIS WITH LC-MS [1]

A highly sensitive ultra-high performance liquid chromatography/tandem mass spectrometry method with in-source fragmentation for rapid quantification of raspberry ketone

[1] This work has been published and refers to: Yuan, B., Zhao, D., Du, R., Kshatriya, D., Bello, N. T., Simon, J. E., & Wu, Q. (2019). A highly sensitive ultra-high performance liquid chromatography/tandem mass spectrometry method with in-source fragmentation for rapid quantification of raspberry ketone. journal of food and drug analysis, 27(3), 778-785.

Abstract

Raspberry ketone (RK) is the characteristic aromatic compound in raspberry (*Rubus idaeus* L.) with wide applications as food additive and anti-obesity agent. However, quantification of RK has presented difficulties in MS detection and reliable LC-MS method for RK analysis in literature is in limit to date. In order to facilitate quality control of raspberry derived products and RK metabolomics study, this study aimed to develop a validated and sensitive UHPLC-MS/MS method. Strong in-source fragmentation was noted and the fragmental ion of 107 *m/z* produced was selected as the precursor ion for MRM detection, and as such the electrospray ionization performance was optimized by fractional factorial design to accommodate such ion-source dissociation behavior as well as its moderate volatility. A pathway involving the formation of quinone-like structure with strong conjugation was proposed to explain the intense in-source fragmentation. The MRM transition was optimized with product ion of 77 *m/z* selected as the quantifier ion. The method featured low limit of quantification of ~2 ng/mL and allowed for rapid detection of RK in fresh raspberries following direct sample preparation. RK contents were found to be higher from locally grown and harvested farm sources compared to commercial products shipped into the state, and higher in those at late-stage compared with early-stage maturity. No correlations in RK content between organic and non-organic labels were noted.

1. Introduction

Red raspberry (*Rubus idaeus* L.) has been a commonly consumed berry fruit for hundreds of years and remains a highly popular fruit. The appeal of red raspberries to consumers largely arises from the berry's characteristic taste and aroma. Among the large number of volatile compounds identified, 4-(4-hydroxyphenyl)-2-butanone, commonly known as raspberry ketone (RK), is recognized as the primary compound responsible for the characteristic raspberry flavor [1-3]. In addition, RK is an FDA-designated generally recognized as safe (GRAS) additive, which has been widely used in the perfumery, cosmetics and food industry to impart raspberry aroma [4]. Due to its low abundance in nature, fruit-derived RK is among the most expensive natural flavor compounds with an estimated market value up to $20,000/kg [5]. Based on the structural similarity to other phenolic compounds (e.g., ephedrine, synephrine, and capsaicin and zingerone), RK also has been investigated as a putative weight loss supplement and appetite suppressant [6]. Rodent studies indicated that RK protected animals from high-fat diet-induced nonalcoholic steatohepatitis [7], prevented diet-induced obesity, and reduced the inclination towards high-fat diets [7, 8]. *In vitro* studies also suggested that RK activates pathways that promotes fatty acid oxidation and reduces lipogenesis in adipocytes [9, 10]. Therefore, having an effective methodology for measuring RK content in raspberries of all sources would better facilitate selection of raspberries that are more appealing to consumers and richer in bioactive components including RK.

The analysis of RK content in red raspberry sources has been predominated by GC-MS as reported in literature [1]. However, there is limited reliable study employing LC-MS methodologies to detect and measure content of RK in red raspberries and related

products. In the study by Urska *et al*. in 2012, a targeted metabolomics method using LC-MS was established for analyzing up to 135 phenolics in fruit with RK included as one of the metabolites. This method, unfortunately, lacked specificity for RK and the limit of detection was not low enough to allow detection of RK in raspberries following direct sample preparation [11].

The aim of this study was to develop a rapid and sensitive method using ultra-high performance liquid chromatography (UHPLC) coupled with triple quadrupole mass spectrometry (QqQ-MS) for reliable quantification of RK in different sources of red raspberries to facilitate quality control, and could also extended the application for RK pharmacokinetic study.

2. Materials and methods

2.1. Chemical reagents and raspberries

Reference standard of RK was purchased from Sigma-Aldrich (St. Louis, MO, USA). Methanol, acetic acid, and HPLC grade water and acetonitrile and formic acid were purchased from Fisher Scientific (Fair Lawn, NJ, USA). Fresh raspberry fruits at different stages of maturity were harvested from local New Jersey farms and stored at -20°C prior to analysis. Fresh raspberries marketed as organic or non-organic products were purchased from local supermarkets, stored at 4°C and then analyzed within two days. The harvest or purchase dates and location, and sample conditions are shown in **Table B-1.**

2.2. Standard and sample preparation

For the standard preparation, approximately 10 mg of RK standard was accurately weighed and dissolved in 25 mL methanol as stock solution. This was further diluted with

70 % methanol for use as the work solution. For sample preparation, frozen fresh fruits of raspberry were first ground with liquid nitrogen, and approximately 3 g was then subsampled, accurately weighed and extracted using 8 mL pure methanol. The mixture was vigorously vortexed for 1 min, sonicated for 5 min and then centrifuged at 3000 rpm for 10 min. The supernatant was transferred to a glass vial and the precipitate was extracted two more times with 8 mL methanol likewise. The supernatants were combined and brought to a final volume of 30 mL. The extract was then diluted 5-fold with 70 % methanol and centrifuged at 13,000 rpm for 10 min prior to LC-MS injection. Three extracts were prepared for each raspberry sample. The final RK content was presented as μg/100 g fresh weight (FW).

Table B-1. Identification of the raspberry samples used in this study and their respective raspberry ketone concentration.

No.	Source/Brand	Purchase/Harvest location	Purchase/ harvest time	Content (μg/kg FW)
1	Driscoll's, mature and red, USDA certified organic	Shoprite, Piscataway, NJ	October, 2017	293.5 ± 40.9
2	Driscoll's, mature and red	Target, Piscataway, NJ	October, 2017	82.0 ± 2.6
3	Driscoll's, mature and red, USDA certified organic	Target, Piscataway, NJ	October, 2017	93.9 ± 3.1
4	Driscoll's, mature and red	Stop & Shop, Piscataway, NJ	October, 2017	74.1 ± 14.5
5	Driscoll's, mature and red, USDA certified organic	Stop & Shop, Piscataway, NJ	October, 2017	296.0 ± 19.7
6	Driscoll's, mature and red, USDA certified organic, (trademark1)	Trader Joe's, New Brunswick, NJ	October, 2017	40.2 ± 1.4
7	Driscoll's, mature and red, USDA certified organic, (trademark 2)	Trader Joe's, New Brunswick, NJ	October, 2017	9.9 ± 2.1
8	Farm berries, half mature, pink to red	Hacklebarney Farms Cider Mill, Chester, NJ	August, 2016	480.9 ± 24.0
9	Farm berries, mature, dark and deep red	Hacklebarney Farms Cider Mill, Chester, NJ	August, 2016	712.1 ± 67.0
10	Farm berries, half mature, pink to red	Rutgers University Cook Organic garden, New Brunswick, NJ	July, 2016	416.4 ± 60.0

11	Farm berries, mature, dark and deep red	Rutgers University Cook Organic garden, New Brunswick, NJ	July, 2016	622.0 ± 44.4

Notes: FW, fresh fruit weight.

2.3. Instrumentation

An Agilent 1100 series LC/MSD instrument (Agilent Technologies, Waldbronn, Germany) was used to facilitate determination of the precursor ion of RK. The HPLC was equipped with an auto-degasser, quaternary pump, column thermostat and a diode-array detector (DAD). Column Phenomenex Luna C18 (2), 150 x 4.60 mm, 5 μm (Torrance, CA) was used for compound separation. The LC-MS interface was electrospray ionization (ESI). Nitrogen was used as nebulizing gas and drying gas. The MS featured an ion trap analyzer and helium was used as the collision gas. Data was acquired using the Agilent ChemStation (ver A.08.03) and LC/MSD Trap Control (ver 5.1).

An Agilent 1290 Infinity II UHPLC coupled with 6470 triple quadrupole (QqQ) (Agilent Technologies, Waldbronn, Germany) was used for development of fully optimized method of quantification of RK. The UHPLC was equipped with a built-in auto-degasser, binary pump and column thermostat. The DAD was bypassed to reduce peak broadening. Waters Acquity BEH C18 column, 50 x 2.1 mm, 1.7 μm (Milford, MA) equipped with Waters Acquity UPLC BEH C8 VanGuard pre-column 5 x 2.1 mm, 1.7 μm (Milford, MA) was used for compound separation. The LC-MS interface was electrospray ionization (ESI) with jet stream. Nitrogen was used as nebulizing gas, drying gas, sheath gas and collision gas. MassHunter Workstation LC/MS Data Acquisition (ver B.08.00) was used for data acquisition and MassHunter Workstation Optimizer (ver B.08.00) for MRM optimization.

2.4. Determination of precursor ion by ion trap MS

Agilent 1100 series ion trap MS was used for screening of the precursor ion. Specifically, for HPLC, mobile phase A was water with 0.1 % formic acid and mobile phase B was acetonitrile with 0.1 % formic acid at a flow rate of 1 mL/min. The gradient started at 30 % B at 0 min and increased to 60 % B at 15 min. The injection volume was 20 μL and around 800 ng of RK was injected onto column. The column thermostat was set at 25 °C. The wavelength of DAD was set at 254 nm with the reference wavelength at 400 nm. About a third of the HPLC eluent was split into MS. For MS, the nebulizer was set at 40 psi, drying gas temperature at 350 °C with a flow rate of 12 L/min. The capillary voltage was +3500 V in positive scan and -3500 V in negative scan. Either positive or negative polarity was used in separate injections. The full scanning range of ion trap was from 50 to 500 *m/z*. Collision energy noted as compound stability was set at 80%, which generally allowed structurally similar small phenolic acids to retain integrity in the ion trap. Ion charge control (ICC) was set with a target of 40,000 with a maximum accumulation time of 300 ms.

2.5. Determination of precursor ion by QqQ-MS

Following the experiment using ion trap MS, Agilent 1290 UHPLC-6470 QqQ was used as the principal instrument in the successive studies. First, full scan mode was used to identify the precursor ion. For UHPLC, the mobile phase components were the same as in the ion trap experiment with a flow rate of 0.4 mL/min. 2. The injection volume was 5 μL and around 5.5 ng standard of RK was injected onto column. The column thermostat was set at 30°C. With respect to QqQ-MS, the ESI was preliminary set as environment suitable

for small phenolic acids with modification, i.e., nebulizer was set at 30 psi, drying gas temperature at 300°C with a flow rate of 12.0 L/min, sheath gas temperature at 300°C with a flow rate of 10.0 L/min, capillary voltage at +2500 V (positive scan) or -2500 V (negative scan), and nozzle voltage at +1000 V (positive scan) or -1000 V (negative scan) [12]**.** Either positive or negative polarity was used in separate injections. Three segments were included in one run, each having the same scan range from 50 to 200 *m/z* and scanning time of 500 ms but different fragmentor voltage (FV) (voltage applied to the exit end of the capillary) at 80, 110, and 140 V, respectively. The three FVs represented a reasonably wide range found to be suitable for structurally similar small phenolic acids [12]. The accuracy of the mass of the predominant ion, the prospective precursor, was further confirmed by full scan using narrower scanning range down to ± 5 Da of the detected mass.

2.6. Optimization of MRM transitions

To facilitate MRM optimization, an isocratic gradient with 28% B was employed and the RK peak was eluted out within 1 min. Other parameters remained the same as those in the precursor confirmation study using QqQ (section 2.5) unless otherwise specialized. The most abundant RK fragment or cluster ions detected in section 2.5 was manually added to the MassHunter Optimizer as the prospective precursor, and then subjected to an optimization procedure composed of five consecutive injections under scanning modes of selected ion monitoring (SIM), SIM, product ion (PI) scan, MRM and then PI, respectively. The injection steps are shown in . Around 5 ng of RK was injected onto column for SIM and MRM and 10 ng injected for PI scan to compensate for the low sensitivity of PI scan

mode. The most abundant product ion was selected as the quantifier ion as the result of a successful optimization.

2.7. Optimization of ESI

The optimized MRM transitions so far acquired and relevant LC-MS conditions as aforementioned were applied for the optimization of ESI. Seven key parameters, i.e., drying gas temperature (X_1) and flow rate (X_2), nebulizer pressure (X_3), sheath gas temperature (X_4) and flow rate (X_5), capillary voltage (X_6) and nozzle voltage (X_7), were optimized by the design of experiment (DOE) approach using fractional factorial design. Two levels of each factor were tested by expanding the general setting to an empirically higher and lower end. As the instrumental sensitivity was dependent on signal response (peak area) and background noise, signal-to-noise ratio (S/N) was used as the DOE model response. Injections were made in triplicate for each trial and the averaged S/N was used as the response. The factorial design conditions are presented in **Table C-2**.

2.8. Optimization of other parameters

For the UHPLC part, influences of mobile phase modifiers, i.e. formic acid and acetic acid at 0.1 or 0.2% on detection sensitivity were studied. For the QqQ MS part, the quadrupole resolution set at either “unit” or “wide” was compared for impact on detection sensitivity. The cell accelerator voltage was studied in a range of 4 to 8 V. The detector Delta EMV was fine-tuned ranging from 0 to 100 V.

2.9. Method validation

The method was validated in terms of linearity range, low limit of detection (LLOD), low limit of quantification (LLOQ), accuracy and intra-batch and inter-batch precision. The LLOD and LLOQ were defined as S/N ratio at 3 and 10, respectively. For accuracy validation, a known amount of RK standard was spiked in the quality control (QC) sample at 200%, 100% and 50% level of expected concentration, and accuracy was calculated as (detected concentration – endogenous concentration)/spiked concentration x 100 %. Fresh raspberries manually harvested from the Rutgers University Cook College Campus Organic garden, New Brunswick, were used as the QC sample. Precision was calculated as the standard deviation of repeated injections in single sequence for intra-batch precision and separated sequences for inter-batch precision at three levels, i.e., LLOQ, middle point of linearity range (MP), and high limit of quantification (HLOQ).

2.10. Data analysis and statistics

Analysis of data acquired from Agilent 1100 LC-MS system was conducted by Agilent Data Analysis (ver 2.2). Analysis of data acquired from Agilent UHPLC-QqQ MS system was performed on MassHunter Workstation Qualitative Analysis (ver B.07.00) and Quantitative Analysis (ver B.07.01). Fractional factorial design was analyzed by Design Expert (ver 8.0.6).

3. Results and discussion

3.1. Method development and validation

3.1.1. ***Determination of precursor ions***

Precursor ions that are protonated, deprotonated and adducted with cations such as sodium and ammonium are the most common precursors formed in the ESI compartment. However, selection of these most common ions as prospective precursors was found in the preliminary RK study failing to generate reliable product ions with strong MS responses. This suggested formation of cluster ions or severe in-source fragmentation in the ESI compartment. In view of the significantly lower sensitivity under full scan mode of triple quadrupole MS compared with ion-trap MS [13], screening for the possible cluster ion(s) and in-source fragment(s) was first conducted using ion-trap MS. As the ion trap MS featured analogous ESI configuration and parameters with those of QqQ MS, ionization behavior of RK observed in the ion trap MS study could provide valuable reference for the subsequent study using QqQ-MS.

The ion trap MS study conducted under positive polarity using low collision energy revealed low abundance of parent ions that were protonated ($[M+H]^+$, 165 *m/z*) and adduction with a sodium ion ($[M+Na]^+$, 187 *m/z*), while the predominant peak was detected at 107 *m/z* (**Figure B-1**). A separate analysis of RK under negative polarity did not generate any noticeable RK peak.

The major challenge in the following QqQ study was the low sensitivity of full scan and that RK peak was barely visible in total ion chromatogram (TIC), as shown in **Figure B-2** inset. Considering that the injected concentration (5.5 ng injected onto column) was already on the high end, injection of higher concentration to improve peak visibility was

avoided to prevent contamination. With reference to the aforementioned ion trap study, the positively charged ion at 107 *m/z* was tentatively extracted and this successfully led to identification of RK peak at 3.8 min. Careful comparison of the mass spectra of RK with that of the background confirmed the actual MS response of the ion at 107 *m/z* and lack of detectable protonated and adducted parent ions (**Figure B-2**). This explained the unsuccessful detection of RK in the preliminary study inappropriately selecting protonated or adducted parent ions as the precursor ions. In addition, negative polarity did not generate significant RK peak on QqQ MS either. Thus, the positive 107 *m/z* ion formed as the result of in-source fragmentation was selected as the precursor ion for the subsequent MRM study using QqQ MS.

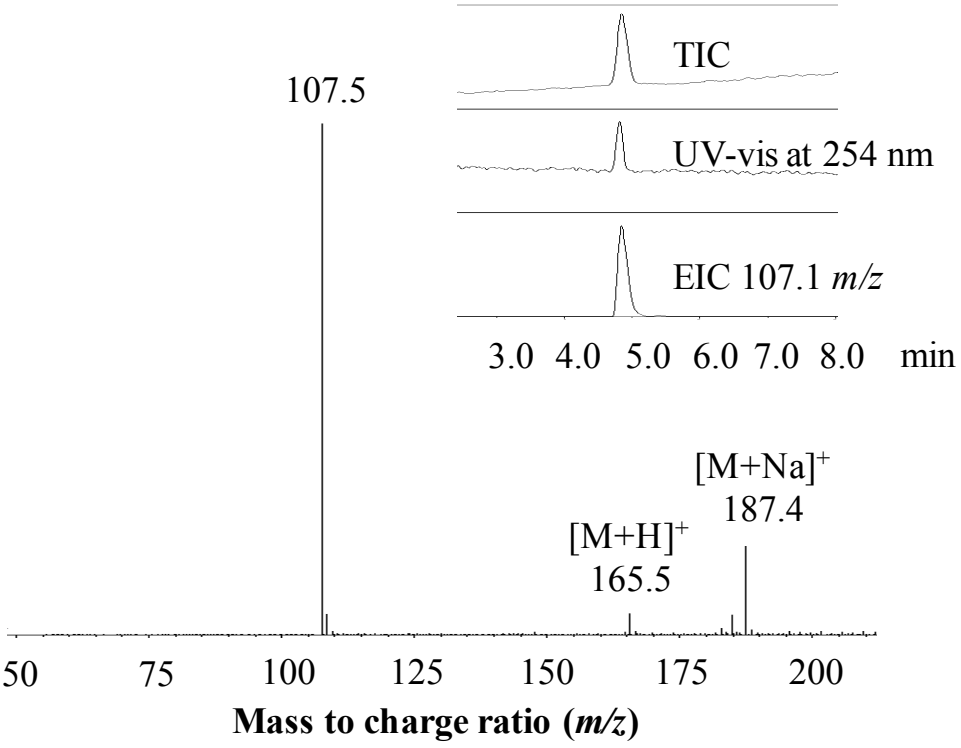


Figure B-1. Mass spectrum of raspberry ketone as acquired by ion trap MS. Inset was the corresponding chromatograms of total ion (TIC), UV-vis at 254 nm and that of extracted ion (EIC) of the major fragment at 107.1 *m/z* with 800 ng raspberry ketone injected on column.

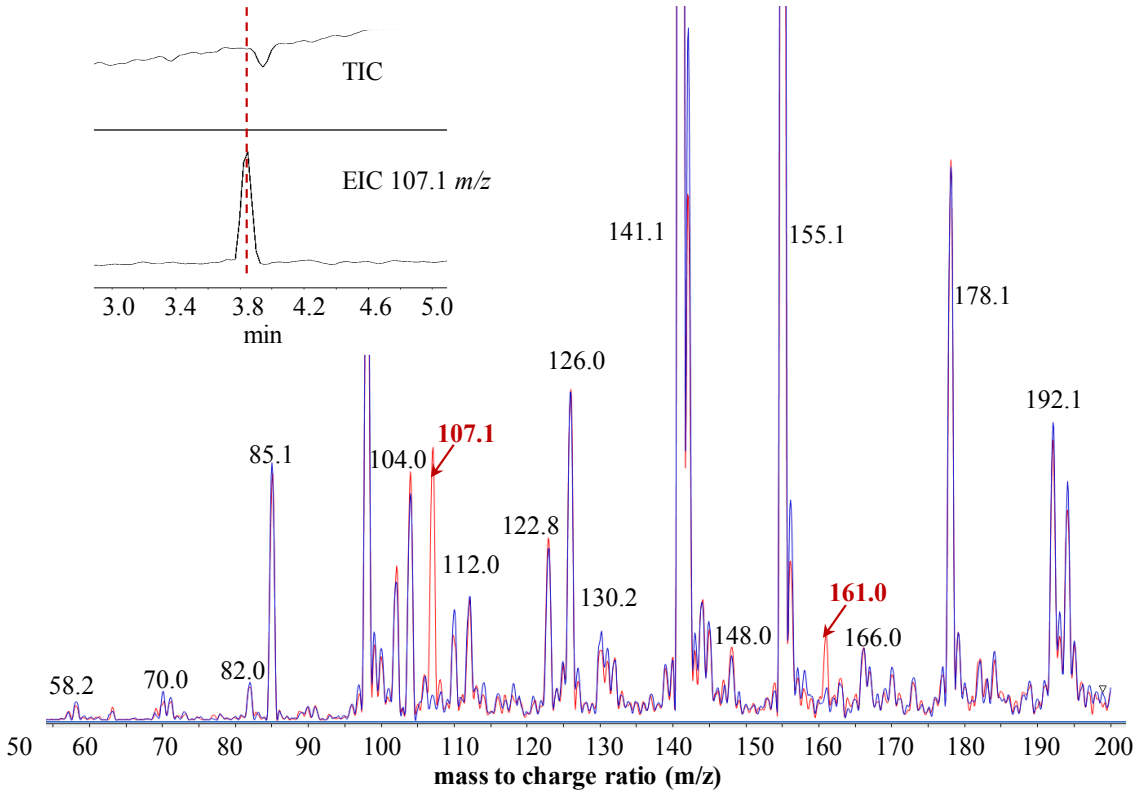


Figure B-2. Overlaid mass spectra of raspberry ketone (RK) and background acquired by QqQ MS. Mass spectrum of RK (red) was acquired at 3.8 min and background (blue) at 4.6 min. Fragmentor voltage was preliminary set at 110 V. Inset in the upper left was the corresponding chromatograms with 5.5 ng of RK injected on column. Notice the near invisibility of RK peak in total ion chromatogram (TIC). RK peak was rendered visible by extracted ion chromatogram (EIC). 161.0 *m/z* was a random ion from the background.

*3.1.2. **Optimization of MRM transitions***

Optimization of MRM transitions can be most conveniently achieved via injection mode instead of the conventional infusion method. Under this injection mode, a connection union can be used in replacement of an analytical column thus without compound separation, so that each injection could be finished within ten seconds. Usage of connection union in the case of RK, however, seemed to result in less accurate "locking" of the characteristic 107 *m/z* ions during the fragmentor voltage (FV)-optimization injections and

led to questionable optimized data. This was realized by comparison of the inaccurately optimized MRM transitions with the corresponding transitions under SIM mode using the same precursor. The inefficiency of optimization was manifested by insignificant improvement or even reduction in S/N acquired under MRM versus SIM mode, as shown in **Figure C-1**. This optimization inadequacy might be a result of interference from isobaric impurities in the injected solvent, which simultaneously entered ESI compartment with RK without column separation.

Thus, an analytical column was used to replace the connection union so as to chromatographically separate RK from possible solvent impurities and to facilitate targeting at the 107 *m/z* ions. Under isocratic elution with 28% B, RK could be timely eluted out at around 1 min with adequate separation from the background impurities, with one optimization cycle (five injections) finished within minutes. The optimization procedure using column was shown in **Figure B-3** and the corresponding key parameters were listed in **Table C-1**. This led to improved optimization efficiency as shown in **Figure C-2**. The eventual optimization result was FV at 130 V and quantifier ion at 77.1 *m/z* under CE 25 eV.

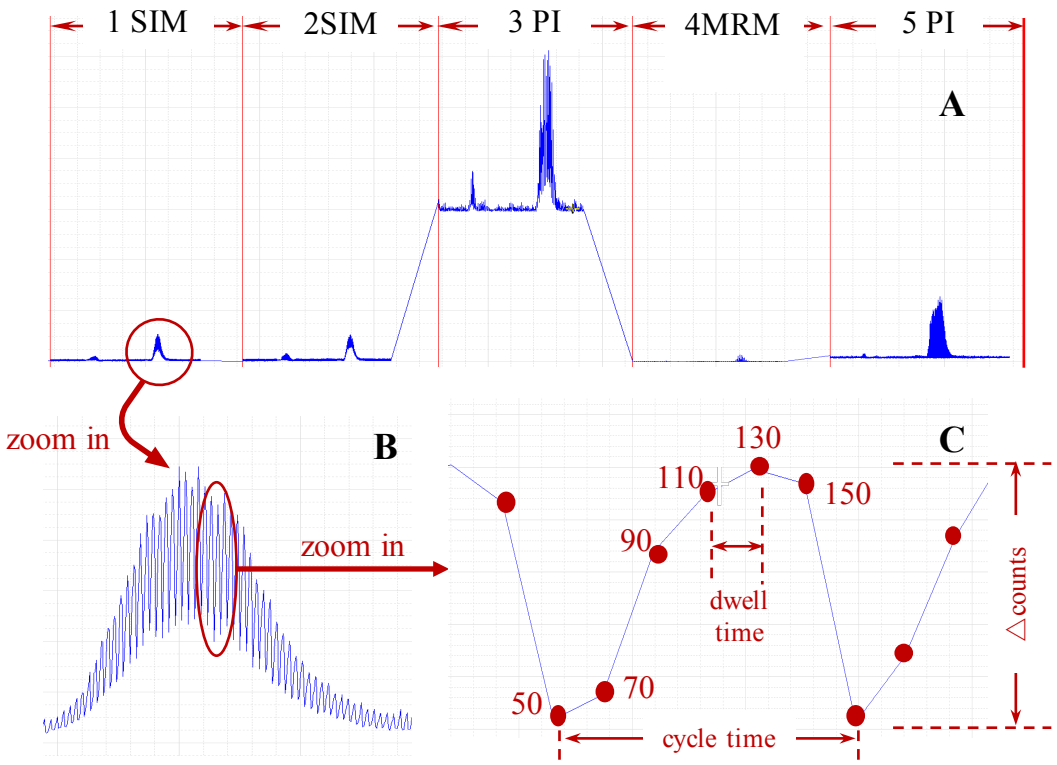


Figure B-3. Real time optimization chromatograms of raspberry ketone (RK). (**A**), chromatographic overview of five consecutive injections and scanning mode; (**B**), zoomed-in peak of RK in the first SIM injection; (**C**) zoomed-in data points (red dots) showing varied signal counts under different fragmentor voltage (numbers). x-axis was time and y -axis signal counts and not shown for clarity. The corresponding settings were shown in **Table C-1**. Notice the 'sawtooth-like' peak curve and "filling" effect under the peak curve due to ∆counts, which indicated differences in sensitivity of signal counts according to the given varying parameter. Also notice the high counts and low S/N in product ion (PI) scan, compared with the low counts and high S/N in MRM.

*3.1.3. **Optimization of ESI***

Unlike most compounds that get ionized remaining intact in the ESI, the intense in-source fragmentation as well as the moderate volatility of RK required the ESI settings to be particularly optimized to achieve the optimal sensitivity. Fractional factorial design was applied to optimize and evaluate the significance of different ESI setting variables and interactions involved, and to reduce trial numbers without losing essential information [14].

The conditions used in the fractional factorial design were included in **Table C-2**. The function using factor codes was established as $Y = -15.56X_1 + 23.94X_2 - 2.69X_3 + 12.44X_4 - 13.69X_5 + 17.31X_6 + 42.81X_7 - 36.06X_1X_2 + 19.56X_1X_3 - 15.81X_2X_4 + 306.69$, $R^2=0.9722$, and adjusted $R^2=0.9166$. The influence and interactions among the seven ESI parameters are schematically depicted in **Figure B-4.** Among the gas-related parameters, S/N was most effectively modulated by changing the temperature and flow rate of drying gas and meantime also fine-tuned by others. When the drying gas flow rate was low, simultaneous elevation in drying and sheath gas temperature and nebulizer pressure significantly increased S/N presumably by increasing the evaporation of solvent in and thus RK desorption from the electrosprayed aerosol. At higher drying gas flow rate, in contrast, optimal S/N could be achieved by reduction in both drying gas temperature and nebulizer pressure. Change in sheath gas temperature had little impact on S/N when the drying gas flow was high. Sheath gas flow was found to be the least important factor with a slight negative impact on S/N. The two voltage settings, nozzle and capillary voltage both had noticeably positive impact on S/N, especially the former being the single most influential in all seven parameters. Higher nozzle voltage beyond the experimented upper bound only generated minor improvement in S/N. Considering all factors collectively, the optimal setting for all seven parameters were determined as drying gas at 250°C with a flow rate at 13 L/min, nebulizer at 25 psi, sheath gas at 300°C with a flow rate at 8 L/min, capillary voltage of 3000 V and nozzle voltage of 1500 V.

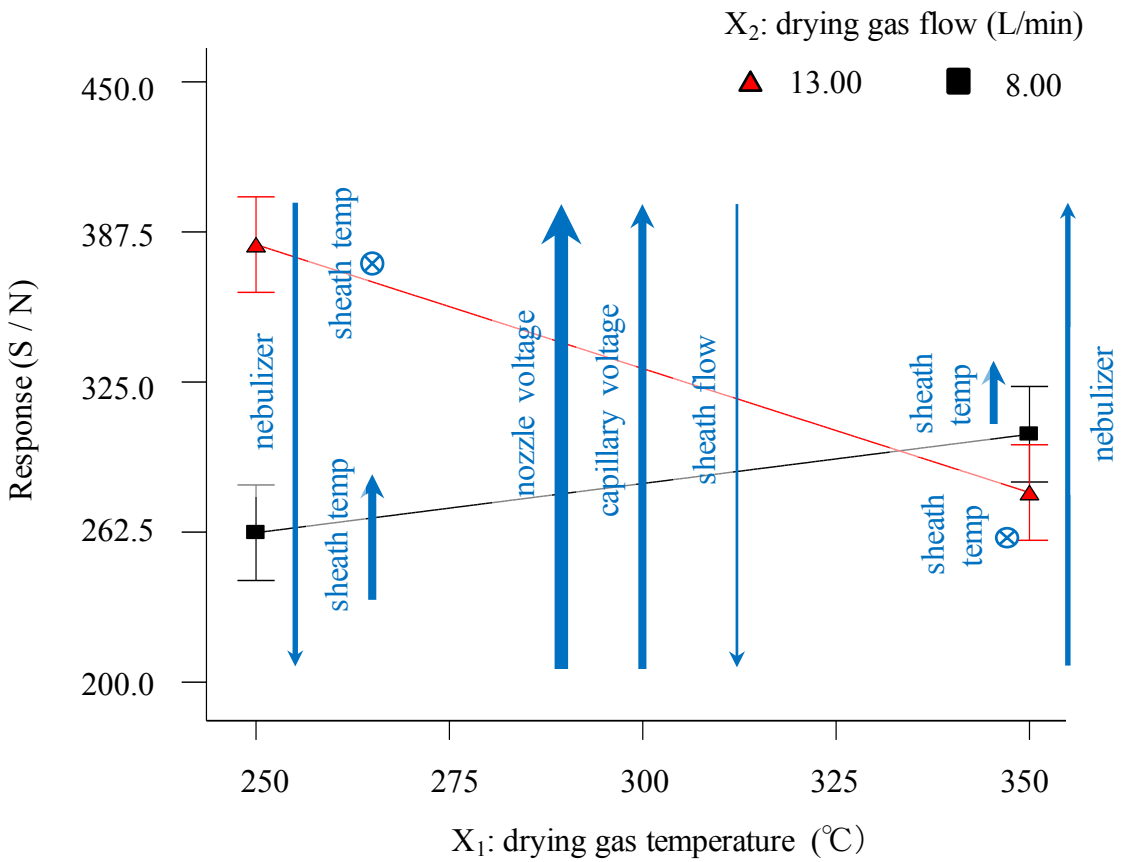


Figure B-4. Impact of ESI settings to signal/noise ratio based on fractional factorial design model. Upward arrows indicate increase in S/N when the corresponding setting increases, and downward arrows indicate increase in S/N when the setting is tuned down. Circled cross indicates negligible influence. Arrows at the plot center indicate direction of vertical gliding of the entire plot when the given variables increase. Arrows at the two sides indicate independent shifting of separated data points. Temp is short for temperature.

3.1.4. Optimization of other settings

The two most commonly used mobile phase modifiers, formic acid and acetic acid, were found to have important impact on instrumental performance. Acetic acid at 0.1% gave the best S/N, and a higher concentration at 0.2% led to marginal reduction in S/N. Addition of formic acid resulted in lower S/N than acetic acid at the same concentrations. Particularly, formic acid at 0.2% led to nearly two times reduction in S/N than 0.2% acetic acid. Mobile phase without modification with formic or acetic acid resulted in high

background and low signal response and thus the lowest S/N. Therefore, 0.1% acetic acid was selected as the optimal mobile phase modifier.

The quadrupole resolution setting defines broadness of the ionic filtration window. Quadrupole resolution set in "unit", filtration window of 0.7 Da wide, was found to give higher S/N than resolution in "wide", which had a filtration window of 1.2 Da wide, by reduction of background noise. Cell accelerator voltage (CAV) was the voltage gradient applied to the collision cell to increase drifting velocity of ions traversing the collision cell and hence to prevent stalling of ions in collision cell and cross-talk between MRMs, the latter being a phenomenon of one ongoing MRM transition getting "contaminated" by product ions produced by the last transition. The range CAV adjusted was from 4 to 8 V, and comparable S/N ratios were obtained. CAV was then set to 5 V. Delta EMV was the extra voltage applied to the detector and adjusted as the last resort for fine tune of sensitivity. S/N increased by around 20% with the elevation of Delta EMV from 10 to 20 V, reached its maximum at around 30 ±10 V, and then decreased slightly and plateaued at higher Delta EMV. Accordingly, the optimal Delta EMV value was set at 30 V.

3.1.5. ***Optimized method and validation***

The optimized method was summarized as below. For the UHPLC part, water with 0.1 % acetic acid was used as mobile phase A, and acetonitrile with 0.1 % acetic acid was used as mobile phase B. The flow rate was 0.4 mL/min. The gradient started at 28% B, held for 1.2 min, then increased to 100% B at 1.3 min and held for 1 min before returning to initial conditions. Eluent between 0 to 0.8 min and after 1.2 min was directed into waste. The column was equilibrated with 28 % B for 1.5 min between injections. The column was

thermostatted at 30 °C. The injection volume was 3 μL. For the QqQ part, the ESI featured a setting of drying gas at 250°C with a flow rate of 13 L/min, nebulizer at 25 psi, sheath gas at 300°C with a flow rate of 8 L/min, capillary voltage of 3000 V and nozzle voltage of 1500 V. The precursor ion was 107 *m/z* with FV at 130 V, and product ions were 77.1 *m/z* with CE of 25 eV as the quantifier ion. Dwell time was 30 ms. The quadruple resolution was "unit". CAV was at 5. Delta EMV value was +30V. Representative MRM chromatograms of standard solution of RK are shown in **Figure C-3**.

The validation result of the method as summarized in **Table B-2** shows excellent accuracy and precision for quantification. The LLOQ was 2 ng/mL or 6 pg injected on column and was low enough to allow for detection of trace level of RK in fresh raspberries following routine analysis. The linearity range had three orders of magnitude, allowing for detection of samples with large dynamic range of content of RK. Accuracies of all QC levels were less than 20% off the expected value. Intra-batch precision was excellent for MP and HLOQ, with all less than 3% deviation, and precision at LLOQ was below 15%. Inter-batch precision as expected showed higher deviation but all below 18% for the three levels validated.

Table B-2. Validation of method for quantification of raspberry ketone.

LLOD (ng/ml)	LLOQ (ng/ml)	Linear range (ng/ml)	Calibration curve	R^2
0.97 ◊	1.95 ◊	1.95 ~ 998.90	Y=33.380 X + 29.634	0.9951 ‡

Accuracy (%)*			Intra-batch precision (%)			Inter-batch precision (%)		
50% level	100% level	200% level	LLOQ	MP	HLOQ	LLOQ	MP	HLOQ
102.7	101.7	116.9	13.8	1.6	2.8	17.6	8.6	8.3

Quant, quantifier ion; LLOD, low limit of detection; LLOQ, low limit of quantification; MP, middle point of linearity range; HLOQ, high limit of quantification. ◊, LLOD and LLOQ were acquired with 3 μL injection volume, corresponding to 2.91 pg and 5.85 pg injected on column, respectively. *, the percentage levels for accuracy refers to the percentage of expected RK concentration in the QC sample. ‡, R^2 was calculated with 1/x weight.

3.2. RK Fragmentation behavior

The fragmentation pathway of RK could be rationalized as shown in **Figure C-4A**. The intense in-source fragmentation of RK could be favored by formation of highly stable fragments with extended conjugation. A possible mechanism started with protonation on the carbonyl site due to its high electronegativity, which triggered electron delocalization for structure rearrangement. This resulted in the cleavage of β bond, a neutral loss of propen-2-ol, and migration of the positive charge to the fragmental ion of 107 *m/z* which had a formula of $C_7H_7O^+$ suggested by high-resolution MS (**Figure C-4B**) with a proposed conjugated quinone-like structure [15, 16]. The precursor ion 107 *m/z* further experienced net loss of CH_2O in the collision cell to form ion 77 *m/z* with formula of $C_6H_5^+$ (**Figure C-4C**) [15]. The positive charge in the product ion was stabilized by the conjugated double bonds. The identities of major fragments were further confirmed by HR-MS.

3.3. Quantification of RK in raspberries

Most published studies related to RK analysis can be dated back to early 1990s, with a broad concentration range reported in raspberries. RK levels were commonly reported in the range of 10 ~700 μg/kg fresh weight (FW) [17, 18], yet one was reported to be up to 4000 μg/kg FW [2]. In this study, fresh raspberries labeled as certified organic or non-organic products were purchased from four different local supermarkets, and cultivated

berries at different stages of maturation were harvested from two local farms. RK was then extracted and quantified. The contents were also found to be divergent, ranging from 10 to 600 μg RK/kg FW, which was within the typical range as previously reported [17, 18]. The RK content levels detected are summarized in **Table B-1**. Representative chromatograms of RK in the raspberry extract are shown in **Figure C-3**. There was a significant difference in RK concentrations among raspberries of the same brand purchased from different supermarkets. While the objective was not to compare total RK relative to what is 'best' but only to ensure the sensitivity of this new method could differentiate the content in different berry sources, the levels of RK did not bear significantly relationship with the screened "organic" or "non-organic" products. There is a significant recognized impact by production and postharvest handling systems, by genetics (e.g. varieties) and seasonal impacts that would impact fruit quality. A strikingly higher level of RK was found in the farm berries than those commercially marketed. As hypothesized, the mature berries contained much higher RK than those only partially mature or artificially ripened.

4. Conclusion

A UHPLC-QqQ/MS method for rapid and sensitive quantification of RK was successfully developed and validated. In particular, full scan experiment using both ion-trap MS and QqQ MS revealed severe in-source fragmentation of RK. Potential mechanism of fragmentation in the ESI and collision cell was proposed for the first time. In view of the unusual in-source fragmentation as well as high volatility of RK, settings for ESI were specially optimized using fractional factorial design, which effectively enhanced the sensitivity. Further, the findings from this study indicate that RK concentration in

commercial and farm berries can be vastly different. This discrepancy could be related to selection process adopted by commercial manufacturers before bringing berries to market. In this regard, organic labeling has little correlation with RK content. A difference in the maturity stage, however, is more likely to influence RK content. Late-stage maturity raspberries were found to have higher RK content, which suggests that RK accumulates with increasing maturity, and this agrees with earlier reports [1]. One limitation of the present study is that there is no sensory evaluation of the red raspberry for taste or aroma. Future work needs to examine the accumulation pattern over a wider degree of fruit maturation. In addition, there is a need to determine influencing factors of RK contents, including light and oxidation, source-sink relationships and other environmental conditions.

5. References

[1] Aprea E, Biasioli F, Gasperi F. Volatile compounds of raspberry fruit: from analytical methods to biological role and sensory impact. Molecules 2015;20:2445-74.

[2] Larsen M, Poll L, Callesen O, Lewis M. Relations between the content of aroma compounds and the sensory evaluation of 10 raspberry varieties (Rubus idaeus L). Acta Agric Scand 1991;41:447-54.

[3] Honkanen E, Pyysalo T, Hirvi T. The aroma of finnish wild raspberries, Rubus idaeus, L. Zeitschrift für Lebensmitteluntersuchung und-Forschung A 1980;171:180-2.

[4] Bredsdorff L, Wedebye EB, Nikolov NG, Hallas-Moller T, Pilegaard K. Raspberry ketone in food supplements--High intake, few toxicity data--A cause for safety concern? Regul Toxicol Pharmacol 2015;73:196-200.

[5] Beekwilder J, van der Meer IM, Sibbesen O, Broekgaarden M, Qvist I, Mikkelsen JD, Hall RD. Microbial production of natural raspberry ketone. Biotechnol J 2007;2:1270-9.

[6] Wang L, Meng X, Zhang F. Raspberry ketone protects rats fed high-fat diets against nonalcoholic steatohepatitis. J Med Food 2012;15:495-503.

[7] Morimoto C, Satoh Y, Hara M, Inoue S, Tsujita T, Okuda H. Anti-obese action of raspberry ketone. Life Sci 2005;77:194-204.

[8] Cotten BM, Diamond SA, Banh T, Hsiao Y-H, Cole RM, Li J, Simons CT, Bruno RS, Belury MA, Vodovotz Y. Raspberry ketone fails to reduce adiposity beyond decreasing food intake in C57BL/6 mice fed a high-fat diet. Food Funct 2017;8:1512-8.

[9] Park KS. Raspberry ketone, a naturally occurring phenolic compound, inhibits adipogenic and lipogenic gene expression in 3T3-L1 adipocytes. Pharm Biol 2015;53:870-5.

[10] Park KS. Raspberry ketone increases both lipolysis and fatty acid oxidation in 3T3-L1 adipocytes. Planta Med 2010;76:1654-8.

[11] Vrhovsek U, Masuero D, Gasperotti M, Franceschi P, Caputi L, Viola R, Mattivi F. A versatile targeted metabolomics method for the rapid quantification of multiple classes of phenolics in fruits and beverages. J Agric Food Chem 2012;60:8831-40.

[12] Zhao D, Bo Yuan; E. Carry; G.M. Pasinetti; L. Ho; J. Faith; I. Mogno; J.E. Simon and Q.L.Wu. . Development and validation of an ultra-high performance liquid chromatography/triple quadrupole tandem mass spectrometry method for analyzing microbial-derived grape polyphenol metabolites. Manuscript in revision. . J Chromatogr B 2018.

[13] Johnson AR, Carlson EE. Collision-Induced Dissociation Mass Spectrometry: A Powerful Tool for Natural Product Structure Elucidation. Anal Chem 2015;87:10668-78.

[14] Martendal E, Budziak D, Carasek E. Application of fractional factorial experimental and Box-Behnken designs for optimization of single-drop microextraction of 2,4,6-trichloroanisole and 2,4,6-tribromoanisole from wine samples. J Chromatogr A 2007;1148:131-6.

[15] ChemCalc: a building block for tomorrow's chemical infrastructure. Patiny, Luc; Borel, Alain Journal of Chemical Information and Modeling 2013. DOI: 10.1021/ci300563h.

[16] Fredenhagen A, Derrien C, Gassmann E. An MS/MS library on an ion-trap instrument for efficient dereplication of natural products. Different fragmentation patterns for [M + H]+ and [M + Na]+ ions. J Nat Prod 2005;68:385-91.

[17] Borejsza-Wysocki W, Goers SK, McArdle RN, Hrazdina G. (p-Hydroxyphenyl) butan-2-one levels in raspberries determined by chromatographic and organoleptic methods. J Agric Food Chem 1992;40:1176-7.

[18] Maquin F, Meili M, Chaveron H. Determination of 4-(p-hydroxyphenyl)-2-butanone by mass fragmentometry. Ann Falsif Expert Chim Toxicol 1981;74:511-21.

APPENDIX C. SUPPORTING MATERIAL OF APPENDIX B

Supporting Material

A highly sensitive ultra-high performance liquid chromatography/tandem mass spectrometry method with in-source fragmentation for rapid quantification of raspberry ketone

Table C-1. Itemized MRM optimization steps for raspberry ketone using MassHunter Optimizer.

1st injection, SIM, 171 ms/cycle, 5.9 cycle/s		
precursor (MS2[§]) (*m/z*)	dwell time (ms)	fragmentor (V)
107.1	25	50
107.1	25	70
107.1	25	90
107.1	25	110
107.1	25	130
107.1	25	150

2nd injection, SIM, 285 ms/cycle, 3.5 cycle/s		
precursor (MS2[§]) (*m/z*)	dwell time (ms)	fragmentor (V)
107.1	25	80
107.1	25	90
107.1	25	100
107.1	25	110
107.1	25	120
107.1	25	130
107.1	25	140
107.1	25	150
107.1	25	160
107.1	25	170

3rd injection, PI, 526.7 ms/cycle, 1.9 cycle/s				
precursor (MS1) (*m/z*)	MS2 scan range (*m/z*)	scan time (ms)	fragmentor (V)	CE (eV)
107.1	50 ~ 174.08	124	130	5
107.1	50 ~ 174.08	124	130	16
107.1	50 ~ 174.08	124	130	27
107.1	50 ~ 174.08	124	130	38

4th injection, MRM, 1111.5 ms/cycle 0.9 cycle/s				
precursor (MS1) (*m/z*)	product ion (MS2) (*m/z*) [‡]	dwell time (ms)	fragmentor (V)	CE (eV)
107.1	109.8 *	25	130	5
107.1	109.8 *	25	130	9
107.1	109.8 *	25	130	13
107.1	109.8 *	25	130	17
107.1	109.8 *	25	130	21
107.1	109.8 *	25	130	25
107.1	109.8 *	25	130	29
107.1	109.8 *	25	130	33
107.1	109.8 *	25	130	37

107.1	109.8 *	25	130	41
107.1	109.8 *	25	130	45
107.1	109.8 *	25	130	49
107.1	109.8 *	25	130	50
107.1	77.1	25	130	5
107.1	77.1	25	130	9
107.1	77.1	25	130	13
107.1	77.1	25	130	17
107.1	77.1	25	130	21
107.1	77.1	25	130	25
107.1	77.1	25	130	29
107.1	77.1	25	130	33
107.1	77.1	25	130	37
107.1	77.1	25	130	41
107.1	77.1	25	130	45
107.1	77.1	25	130	49
107.1	77.1	25	130	50
107.1	51.1†	25	130	5
107.1	51.1†	25	130	9
107.1	51.1†	25	130	13
107.1	51.1†	25	130	17
107.1	51.1†	25	130	21
107.1	51.1†	25	130	25
107.1	51.1†	25	130	29
107.1	51.1†	25	130	33
107.1	51.1†	25	130	37
107.1	51.1†	25	130	41
107.1	51.1†	25	130	45
107.1	51.1†	25	130	49
107.1	51.1†	25	130	50

5th injection, PI, 150 ms/cycle, 6.7 cycle/s

precursor (MS1) (*m/z*)	MS2 scan range (*m/z*)	scan time (ms)	fragmentor (V)	CE (eV)
107.1	49.7 ~ 52.7	50	130	37
107.1	75.6 ~ 78.6	50	130	25
107.1	108.3 ~ 111.3	50	130	5

Notes: SIM, selected ion monitoring; PI, product ion scan; MRM, multiple reaction monitoring; CE, collision energy. §, the QqQ used in this study was composed in sequence of a quadrupole (MS1, the first "Q"), a hexapole which functioned as the collision cell (the small "q". The hexapole traps ions with higher efficiency than quadrupole and thus installed as the collision cell in the QqQ in replacement of a quadrupole), and the second quadrupole (MS2, the second "Q"). In SIM, it was the second quadrupole MS2 that worked as the essential mass analyzer performing ion filtration while MS1 only functioned for ion transmission. ‡, only three product ions out of the designated maximum of four were

detected and picked up during the PI scan as the most abundant and consistent product ions. *, a random wrong "product ion" picked up from background interference. †, 51.1 *m/z* was detected as a qualifier candidate of RK, but was not detected in another independent study by high-resolution Fourier transform MS, and was thus not included in the finalized method due to such inconsistency. Around 5 ng RK was injected on column for SIM and MRM and 10 ng injected for PI scan.

Table C-2. Fractional factorial design conditions for the raspberry ketone studies and model results

Experimental condition and results										Modeled response	
Run order	Standard order	X_1	X_2	X_3	X_4	X_5	X_6	X_7	Response	Predicted	error (%)
1	5	-1	-1	1	-1	1	1	1	237	258	-8.9
2	12	1	1	-1	1	-1	-1	-1	208	212	-2.1
3	9	-1	-1	-1	1	-1	1	1	385	387	-0.4
4	16	1	1	1	1	1	1	1	357	339	5.1
5	3	-1	1	-1	-1	1	1	-1	371	369	0.6
6	15	-1	1	1	1	-1	1	-1	333	345	-3.5
7	14	1	-1	1	1	-1	-1	1	373	388	-3.9
8	8	1	1	1	-1	1	-1	-1	207	225	-8.9
9	2	1	-1	-1	-1	1	-1	1	275	270	1.8
10	6	1	-1	1	-1	-1	1	-1	295	280	5.1
11	10	1	-1	-1	1	1	1	-1	270	275	-2.0
12	7	-1	1	1	-1	-1	-1	1	414	403	2.8
13	13	-1	-1	1	1	1	-1	-1	216	194	10.0
14	11	-1	1	-1	1	1	-1	1	411	413	-0.5
15	1	-1	-1	-1	-1	-1	-1	-1	211	210	0.6
16	4	1	1	-1	-1	-1	1	1	344	339	1.4

Codes for factors and levels

Level	X_1	X_2	X_3	X_4	X_5	X_6	X_7
-1	250	8	25	250	8	1500	500
1	350	13	40	350	12	3000	1500

Signal to noise ratio was used as the model response.

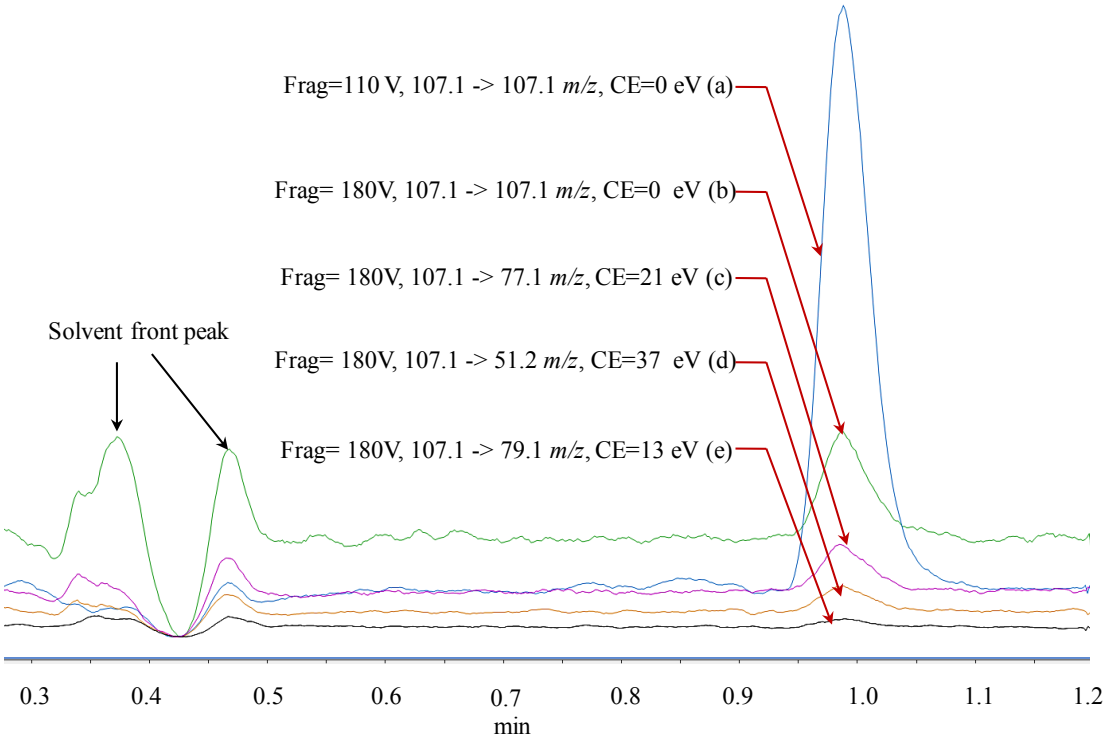


Figure C-1. Inefficient MRM optimization using connection union as shown by overlaid chromatograms under MRM (c, d, and e) and SIM (a and b) modes. Analytical column with 28 % B isocratic elution was used in this experiment to show the background and potential contaminant peak interference. The SIMs were literally run in MRM mode with the product ion *m/z* being the same as that of the precursor with 0 eV collision energy. MRMs of c, d and e showed insignificant improvement in S/N than SIM of b, and SIM b (Frag=180 V, FV optimized incorrectly) featured higher background noise and lower peak height than SIM a (Frag=110 V, a voltage empirically selected also with reference to literature). This suggested an optimization result of either inappropriate fragmentor voltage or wrong product ions. Notice the huge solvent front peak especially on the green line, which suggested that impurities from the background might be a source of interference for optimization without column. Frag is short for fragmentor voltage.

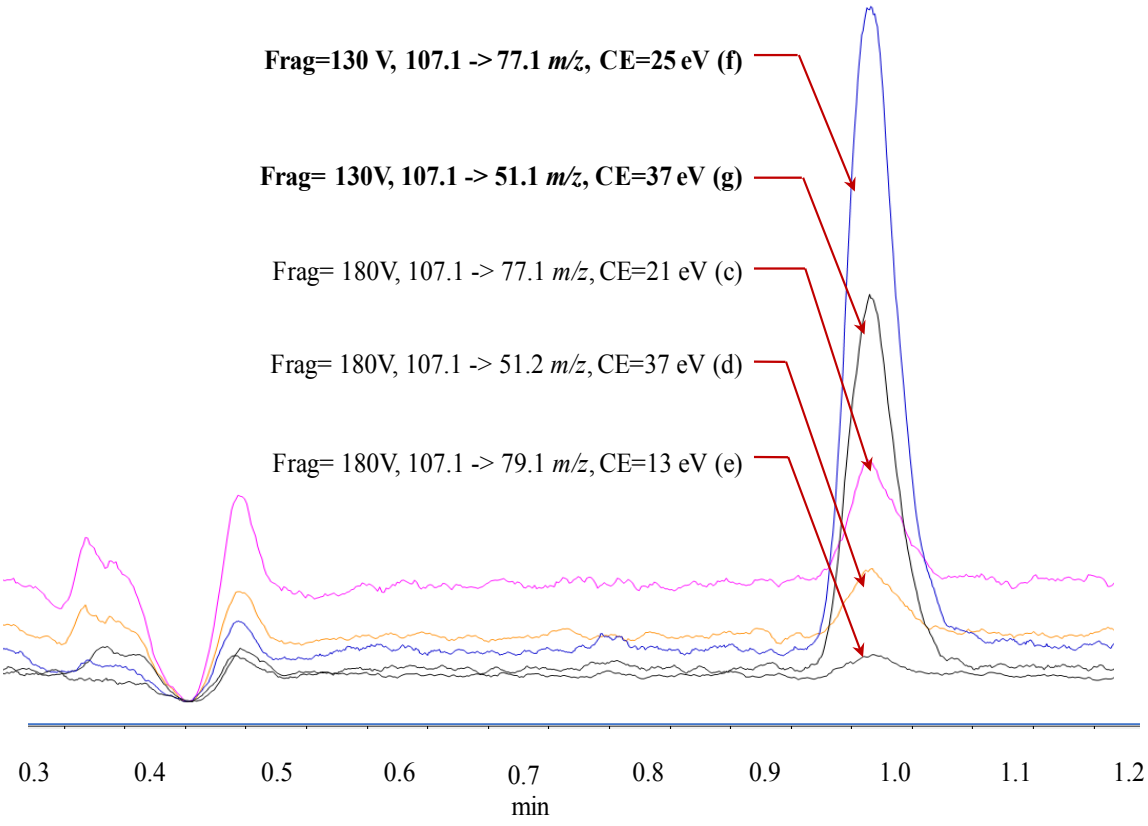


Figure C-2. Overlaid MRM chromatograms under settings optimized by connection union (c, d, and e) and analytical column (f and g, in bold). The improvement using analytical column vs. connection union was manifested by the elevated S/N, featuring both increased peak height and lowered background noise. This improvement was achieved by correction of the fragmentor voltage from 180 V to 130 V. Products 77.1 and 51.2 *m/z* in transitions c and d, respectively, were both potentially valid product ion candidates, but 79.1 *m/z* in transition e was incorrect. It should be noted that the product ion candidate 51 *m/z* was not detected in another independent study by high-resolution Fourier transform MS, and was thus not included in the finalized method due to such inconsistency. Frag is short for fragmentor voltage.

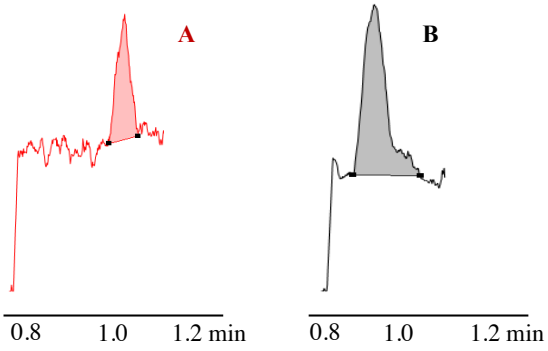


Figure C-3. Representative MRM chromatograms of standard solution of raspberry ketone with 12 pg injected on column (**A**) and extract of mature raspberries manually harvested from Rutgers University Cook Organic garden, New Brunswick, NJ (**B**).

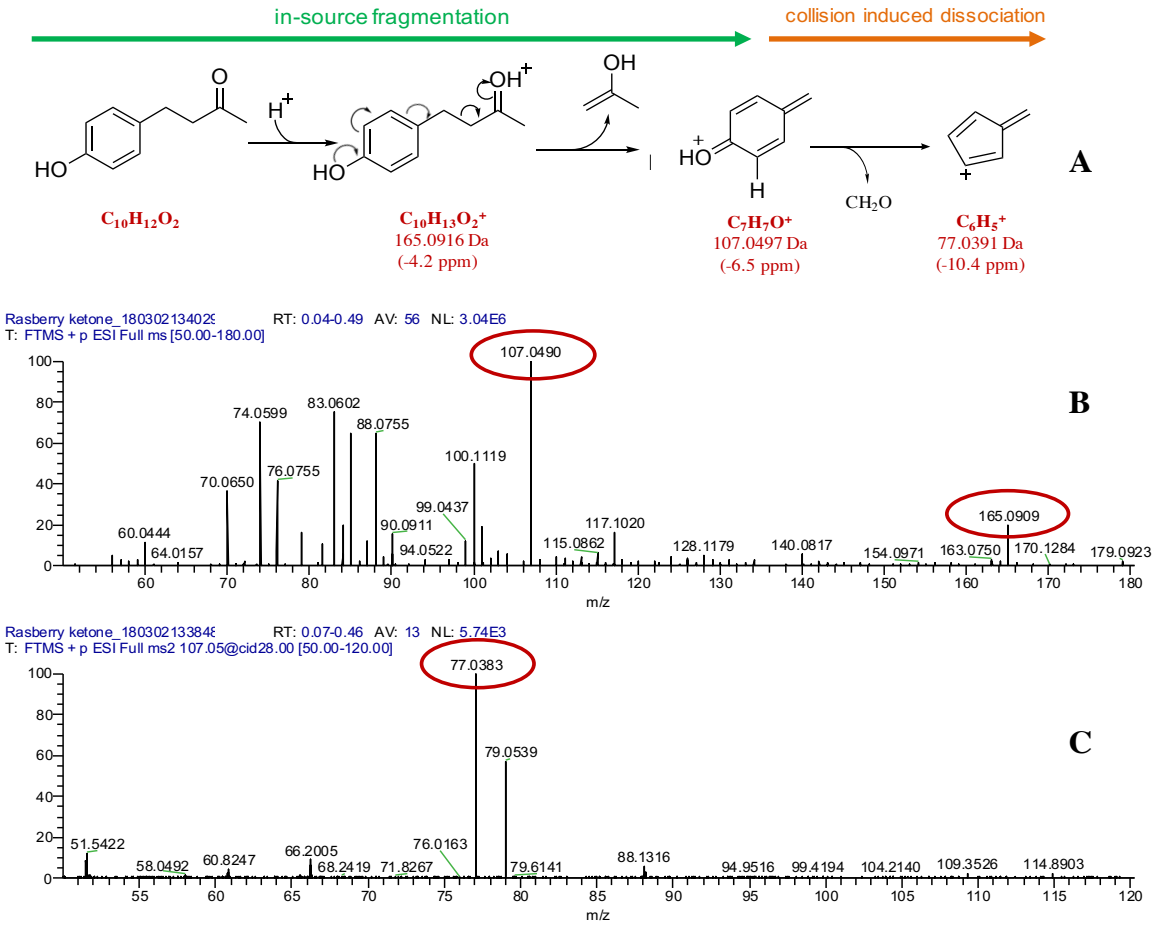


Figure C-4. Raspberry ketone (RK) fragmentation pathway proposed and accurate mass measurement by Fourier-transform mass spectrometry. A, the proposed fragmentation pathway of RK. B, mass spectrum acquired by full scan (50 ~ 180 *m/z*). C, mass spectrum acquired by product ion scan with 107 *m/z* as the precursor and collision energy of 28 eV. Mass accuracy (ppm) was calculated as $(m_i - m_a)/m \times 10^6$, with m_i being the measured mass and m_a the calculated or accurate mass.

APPENDIX D. RASPBERRY KETONE METABOLOMICS

UHPLC-QqQ-MS/MS method development and validation with statistical analysis: determination of raspberry ketone metabolites in mice plasma and brain

Abstract

Raspberry ketone (RK) (4-(4-hydroxyphenyl)-2-butanone) is the major compound responsible for the characteristic aroma of red raspberries, and has long been used commercially as a flavoring agent and recently as a weight loss supplement. A targeted UHPLC-QqQ-MS/MS method was developed and validated for analysis of RK and 25 associated metabolites in mouse plasma and brain. Dispersion and projection analysis and central composite design were used for method optimization. Random effect analysis of variance was applied for validation inference and variation partition. Within this framework, repeatability, a broader sense of precision, was calculated as fraction of accuracy variance, reflecting instrumental imprecision, compound degradation and carry-over effects. Multivariate correlation analysis and principle component analysis were conducted, revealing underlying association among the manifold of method traits. R programming was engaged in streamlined statistical analysis and data visualization. Two particular phenomena, the analytes' background existence in the enzyme solution used for phase II metabolites deconjugation, and the noted liability of analytes in pure solvent at 4 ℃ vs. elevated stability in biomatrices, were found critical to method development and validation. The approach for the method development and validation provided a foundation for experiments that examine RK metabolism and bioavailability.

1. Introduction

4-(4-hydroxyphenyl)-butan-2-one or raspberry ketone (RK) is the major aromatic compound responsible for the characteristic flavor of red raspberries (*Rubus idaeus*) [1], and has been widely used for long time as a generally recognized as safe (GRAS) flavoring substance in foods, perfumery and cosmetics. Due to the low natural concentration (0.01~4 mg/kg fresh weight in raspberries), RK is mainly produced by chemical synthesis.

In recent years, RK has received growing attention for its potential health benefits. RK was shown to reduce lipid accumulation in adipocytes [10, 19], prevent high-fat diet induced obesity in mice [7] and associated nonalcoholic steatohepatitis in rats [6]; and also alleviate ovariectomy-induced obesity in rats [19]. Apart from anti-obesity related benefits, other functions have also been reported, including antiandrogenic activity in the human breast cancer cells [20]; depigmentation activities for zebrafish and mice [21]; anti-inflammatory properties in *E.coli* lipopolysaccharide-stimulated macrophages [22]; as well as cardioprotective action against isoproterenol-induced cardiotoxicity in rats [23]. Of the many health benefits reported, the anti-obesity effects have attracted most attention, and there has been an increased demand for RK as a food supplement for weight loss in recent years, despite a lack of pharmacokinetics and toxicological data. As such, it is imperative to examine the *in vivo* bioavailability and toxicity to ensure safe human consumption at the labeled doses in commercial products [4, 24].

In contrast to the abundant number of studies reporting RK's biological effects, studies on RK metabolism are scarce. In a pioneering work conducted in the 1980s, Sporstøl et al. studied RK metabolites in the urine of rats, guinea-pigs and rabbits using GC/MS [25]. After enzymatic deconjugation, 13 metabolites were identified, with the

reduction product 4-(4-hydroxyphenyl) butan-2-ol or raspberry alcohol (ROH) being the predominant one; other metabolites were derived from modification of RK side chain and/or aromatic ring through hydroxylation, methylation, carboxylation and/or decarboxylation. Apart from this landmark study, there has been no other related study reported.

To comprehensively study RK metabolism, pharmacokinetics and bioavailability, and to prepare for translation into clinical trials to assess RK safety / toxicity at the recommended intake amount in RK-enriched supplements, this work focused on development and validation of an ultra-high performance liquid chromatography with a triple quadrupole tandem mass spectrometry (UHPLC-QqQ-MS/MS) method for targeted determination of RK and its metabolites in plasma and brain specimens from mice. In this work, RK and 25 associated metabolites were investigated which were selected based on the potential RK biopathway(s) and structural similarity to RK [26]. Building upon modern MS/MS methodology, improved statistics and visualization tools with streamlined analysis using R programming were applied for method performance evaluation and validation results interpretation.

2. Materials and Methods

2.1. Chemicals and reagents

Standards of analytes (analytical or reference grade) used included RK (**1**), and its *phenolic aldehyde derivatives*, 4-(4-methoxyphenyl)-2-butanone (RK-Me) or anisylacetone (**2**), benzylideneacetone (PhLiAce) (**3**), 3, 4-dihydroxybenzylideneacetone (3,4-DHPhLiAce) (**4**), vanillylacetone (VLAce) (**5**), vanillylidenacetone (VLiAce) (**6**);

phenolic alcohol derivatives, ROH (**7**), 2-(4-hydroxyphenyl) ethanol or tyrosol (4-HPE) (**8**), 2-(3, 4-dihydroxyphenyl) ethanol (3, 4-DHPE) or 3-hydroxytyrosol (**9**), 4-hydroxybenzyl alcohol (4-HBOH) (**10**); *phenylpropionic derivatives*, 3-(3-hydroxyphenyl) propionic acid (3-HPPA) (**11**), 3-(4-hydroxyphenyl) propionic acid (4-HPPA) (**12**), 3-(3, 4-dihydroxyphenyl) propionic acid (3, 4-DHPPA) (**13**), 3-(3-methoxy, 4-hydroxyphenyl) propionic acid or dihydroferulic acid (DFA) (**14**); *cinnamic acid derivatives*, 4-hydroxycinnamic acid (4-HCA) or *p*-coumaric acid (**15**), ferulic acid (FA) (**16**), caffeic acid (CA) (**17**); *phenyl acetic derivatives*, 3-hydroxyphenylacetic acid (3-HPAA) (**18**), 4-hydroxyphenylacetic acid (4-HPAA) (**19**), 3, 4-dihydroxyphenylacetic acid (3, 4-DHPAA) (**20**); *benzoic acid derivatives*, 3-hydroxybenzoic acid (3-HBA) (**21**), 4-hydroxybenzoic acid (4-HBA) (**22**), 3, 4-dihydroxybenzoic acid or protocatechuic (3, 4-DHBA) (**23**), vanillic acid (VA) (**24**), homovanillic acid (HVA) (**25**), and hippuric acid (HA) (**26**). In addition, *trans*-cinnamic acid-d_7 (**27**) and 4-hydroxybenzoic-d_4 acid (**28**) were used as internal standards (ISs). The chemical structures are presented in **Figure D-1**. The aforementioned standards, and ascorbic acid and β-glucuronidase (from limpets (*Patella vulgate*), ≥85,000 units/mL in contamination with sulfatase) were purchased from Sigma-Aldrich (St. Louis, MO), except that standards of **4** and **14** from Alfa Aesar (Tewksbury, MA), **7** from USP (Rockville, MD), **16** and **23** from ChromaDex (Irvine, CA). Other reagents including methanol, ethyl acetate, glacial acetic acid, formic acid, concentrated hydrochloric acid and LC/MS grade water and acetonitrile were obtained from Fisher Scientific (Pittsburgh, PA). Associated reagent solution preparation for various purposes refers to the supplementary information.

2.2. Instrument

Analytical work was performed using an Agilent 1290 Infinity II UHPLC coupled with 6470 triple quadrupole mass spectrometry (QqQ-MS/MS) with an electrospray ionization (ESI) source (Agilent Technologies, Santa Clara, CA). Nitrogen from a Parker Balston NitroFlow60NA nitrogen generator (Lancaster, NY) was used as the nebulizer gas and collision gas. Chromatographic separation was achieved using a Waters Acquity UPLC BEH C18 column (2.1 × 50 mm, 1.7 μm) with a VanGuard Acquity C18 guard column (2.1 × 5 mm, 1.7 μm) (Milford, MA).

2.3. Mice plasma and brain collection

Seven-week old male C57BL/6J mice (Jackson Laboratory, Bar Harbor, ME) fed on polyphenol-free diet were deeply anesthetized with isoflurane 5% with oxygen for blood collection by cardiac puncture. Plasma was acquired after blood centrifugation at 3000 ×g for 10 min at 4 °C, and then acidified with 2% formic acid to a final concentration of 0.2% (v/v). After cardiac puncture and exsanguination, and perfusion with 0.9% saline, brains were excised, homogenized with 0.2% formic acid (1:2, w/v) and snap frozen in liquid nitrogen. Plasma and brain were stored at -80 °C before analysis. All protocols involving animals were approved by the Institutional animal Care and Use Committee of Rutgers University (OLAW #A3262-01, protocol #13-001).

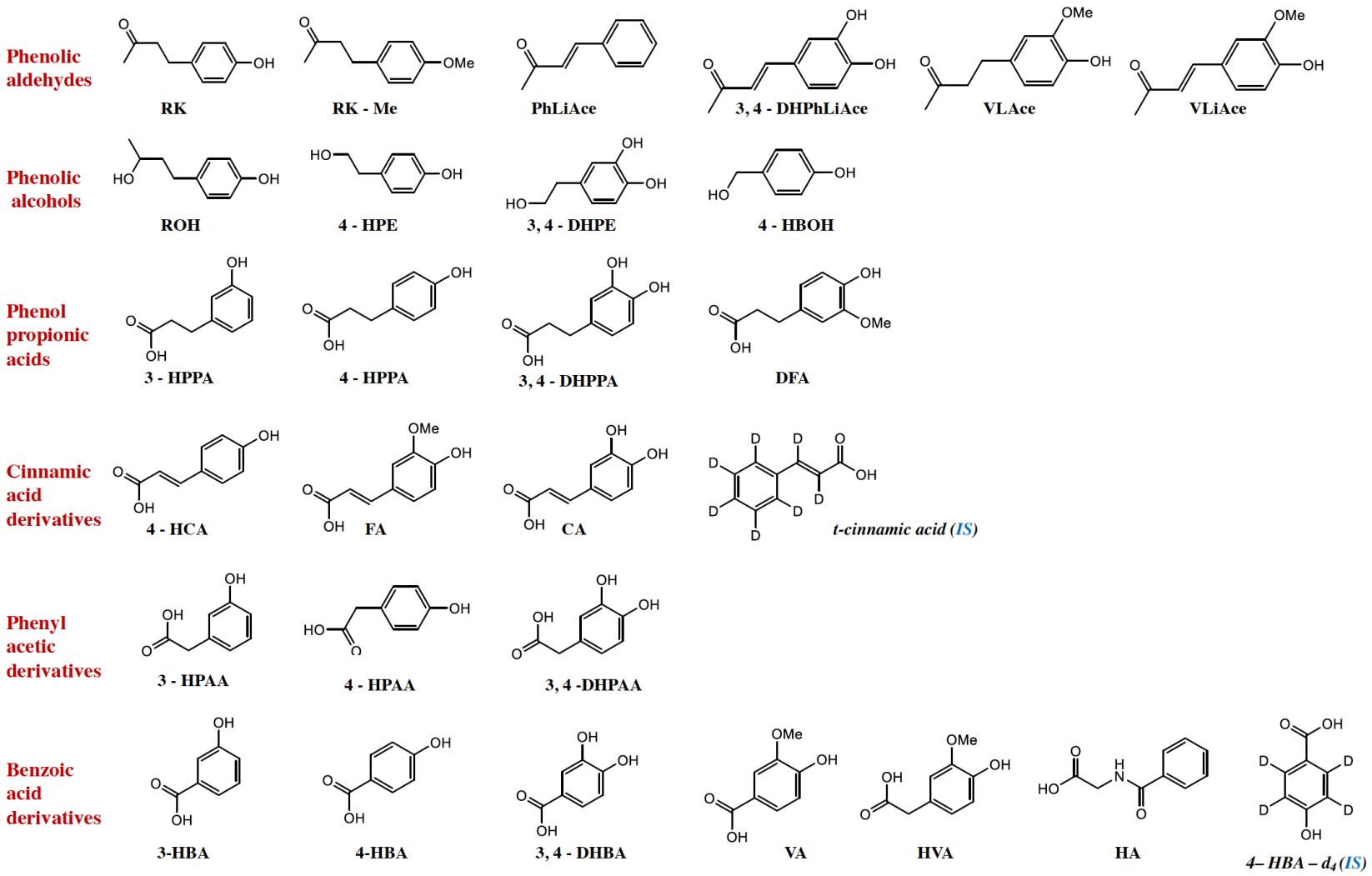


Figure D-1. Analytes' structure, categories and abbreviations. Abbreviations refer to section 2.1.

2.4. Sample preparation

For preparation of reference standard solution, about 15 mg of each standard was accurately weighted and prepared in 25 mL 70% methanol with 0.1% formic acid as stock solution, as then separately aliquoted into 1.5 mL Eppendorf tubes and stored under -80℃. Stock solutions for each compound (except internal standards (IS)) after conditioned to room temperature were mixed as a standard cocktail, and then diluted with 60% methanol with 0.1% formic acid to desired concentration (~200 ng/mL) for instrumental optimization, or diluted with the same solvent into serial concentration (0.1 ng/mL ~ 6 μg/mL) with spiked IS (~100 ng/mL) for calibration. For β-glucuronidase solution (~2000 U) preparation, the original enzyme extract was diluted by 40 times using NaH_2PO_4 buffer (0.4 mol/L, pH 5.0).

For analyte extraction from biomatrices, an 100 μL aliquot of plasma was thawed on ice followed by adding 5 μL of each IS solution (*ca* 2 μg/mL), 300 μL of 0.4 M NaH_2PO_4 buffer (pH 5.4), and 50 μL of β-glucuronidase solution (~2000 U diluted in NaH_2PO_4 buffer). The cocktail was gently mixed, briefly purged with nitrogen to exclude headspace oxygen, and then incubated at 37°C for 45 minutes. The analytes were then extracted with 500 μL of ethyl acetate, vigorously vortexed for 10 sec, sonicated in ice water for 10 min, and then centrifuged at 5000 ×g for 5 min. The supernatant was collected in a glass tube containing 20 μL 2% ascorbic acid methanol solution. The precipitate was then extracted in like manner for two more times. The pooled supernatant was dried under a gentle stream of nitrogen. The residue was reconstituted in 100 μL of 60% methanol containing 0.1% formic acid, centrifuged at 16, 000 ×g for 10 min before LC-MS analysis. The brain samples were processed in similar procedure as plasma, except the following: the tissue

amount used was 500 μL; enzyme solution amount used was 100 μL; after incubation, 100 μL of 4% HCl was added before extraction to denature and precipitate proteins.

2.5. UHPLC-QqQ-MS/MS method

For chromatographic separation, water with 0.1% acetic acid (AA) was used as mobile phase A and acetonitrile with 0.1% AA as phase B, with a flow rate at 0.45 mL/min. The gradient elution (noted as B%) was 5% at 0 min; 10% at 0.5 min; 28% at 3.8 min; 40% at 3.9 min; 55% at 5.5 min; 80% at 5.6 min and then held isocratically until 6 min. The column was equilibrated for 2.5 min before next injection. The column was thermostatted at 30 °C and the autosampler maintained at 4°C. The injection volume was 3.5 μL.

For MS analysis, a further statistical analysis was conducted upon prior reported RK ESI 2_{IV}^{7-3} fractional factorial design [27] as preparation for ESI optimization of all other 25 analytes. Dispersion analysis was conducted to investigate instrumental stability operated at each parameter level, and projection analysis was performed to select and confirm important ESI parameters for further optimization [28] . Following that, drying gas temperature (DGT), drying gas flow rate (DGF) and nozzle voltage (NV), confirmed as the most important ESI parameters, were then further tuned for all analytes, particularly with DGT and DGF optimized collectively by central composite design (CCD). The final ESI conditions were set at DGT 200℃, DGF 12 L/min, and NV +1500 / -1000 V; as to other ESI settings, nebulizer pressure at 30 psi, sheath gas temperature at 250 ℃ with its flow rate at 8 L/min, and capillary voltage at + 3000 V/ - 2500 V. The MS was operated in dynamic multiple reaction monitoring (dMRM) mode with switching polarities, optimized as previously described [27, 29].

2.6. Method validation

The validation procedure followed U. S. Food and Drug Administration guidelines and relevant literature [30, 31] with necessary adaptation. For accuracy assessment, quality control samples (QCS's) were prepared by spiking blank biomatrices with standard mixture containing all analytes at four levels (A, 2000 ng/mL; B, 1000 ng/mL; C, 150 ng/mL; and D, 15 ng/mL, concentration in final processed samples to be injected), each level with five replicates. All QCS's were injected in *randomized* order, with duplicate injections spaced by *ca* 10 hours in a single sequence as a simulation of a typical batch time. Accuracy was computed following the rule of error propagation and random effects analysis of variance (RND-ANOVA). Repeatability was calculated as the mean square error associated with RND-ANOVA variance partition. Validation of matrix effects, recovery and processing efficiency, adapted from the approach by Matuszewski et al. [32, 33], comprised two-level (B, C) spiking *post*-extraction (vs. spiking *pre*-extraction for accuracy validation) and spiking in pure solvent, with calculation following the error propagation rule. Method validation results and associated statistical quantities were then subjected to multivariate correlation analysis and principle component analysis (PCA) [34]. Associated formulas are shown in Supplementary Material.

2.7. Statistics analysis

Microsoft Excel (version 16.16.5), Design Expert (version 8.0.6) and R (version 1.1.463) were used for statistical computation [35, 36]. The R script constructed for data

analysis refers to https://yuanbofaith.github.io/RK_LCMS/. The original data from which the script reads refer to the Supplementary Material.

3. Results and Discussion

3.1. ESI dispersion and projection analysis

Dispersion analysis upon a prior reported RK-oriented ESI 2_{IV}^{7-3} fractional factorial design [27] revealed in this work higher measurement volatility at the elevated level of nebulizer pressure as well as sheath gas flow, and thus the lower levels for both settings were used in the developed method of this work. The magnitude of other ESI settings did not exert noticeable impact on performance consistency (**Figure E-1**). In addition, the prior work reported the large effects of DGT, DGF and NV, yet without considering what is known as the alias structure, i.e., the apparent effects of investigated factors were in fact confounded or "contaminated" with other effects (see Supplementary spreadsheet "Fractional factorial"). To clear-up the alias effects, projection analysis was conducted in this work by collapsing the original design into two replicates of 2^3 full factorial design of DGT, DGF and NV (**Table E-1**, **Table E-2** and **Table E-3**) while treating other ESI factors as background noise, and indicated more than 70% accountability for total data variability from the three factors alone. As such, the three factors were subjected to further optimization for all metabolites investigated.

3.2. DGT and DGF optimization by CCD

As DGT and DGF presented strong interaction while negligible interaction with NV as suggested by projection analysis, DGT and DGF were collectively optimized using CCD

(unlike NV tuned independently), with CCD design displayed in **Table E-4**. A quadratic model was used to approximate signal responses of all analytes, shown in **Figure D-2-A-C**. Generally, higher signal response was favored by increased DGF, and thus 12 L/min was selected as the final DGF. For DGT, special consideration was given to VLAce and 3-HPAA, both of which showed highest susceptibility to DGT but in an opposite manner, i.e., one was favored at low level while the other at high end. As such, 200℃ was selected as the DGT. Modelling efficiency was strongly associated with measurement consistency, which was manifested by the degree of scattering of CCD center points (repetitive measurements at the middle level of the tested factors), as shown in **Figure E-2**. Compound degradation (**Figure E-3**) over the period of CCD batch time was later found to be the cause of measurement inconsistency, accounting for 75% of modelling inadequacy, as shown in **Figure D-2-D**.

3.3. dMRM transitions

The dMRM parameters are displayed in **Table D-1**. Generally, phenolic acids showed higher sensitivity under negative than positive polarity by easy deprotonation of the carboxylic group. Most product ions were formed by subsequent loss of the carboxyl group by 44 Da, in agreement with prior research [37]. Other product ions were generated by loss of a methyl group for the precursors with methoxy group (e.g., VA, 167 -> 152 *m/z*; FA, 193 -> 178 *m/z*), cleavage of a phenyl bond (e.g., HA, 178 -> 77 *m/z*) or rupture of the aromatic ring (e.g., 3-HPAA, 151 - > 65 *m/z*). Phenolic aldehydes and alcohols generally exhibited higher sensitivity under positive polarity, and in-source fragmentation was noticeable for many such compounds. RK and its respective reduced and methylated

derivatives ROH and RK-Me, for example, had intense in-source fragmentation by cleavage of the beta-bond (or the equivalent benzyl bond) [27], and 4-HPE featured in-source dehydration. Fragmental ions produced in the ESI chamber if predominant were selected as the precursor ions for MRM transitions. Such in-source phenomenon rendered analogues RK, ROH and 4-HBOH spectrometrically undistinguishable by having identical MRMs, but they were efficiently resolved chromatographically.

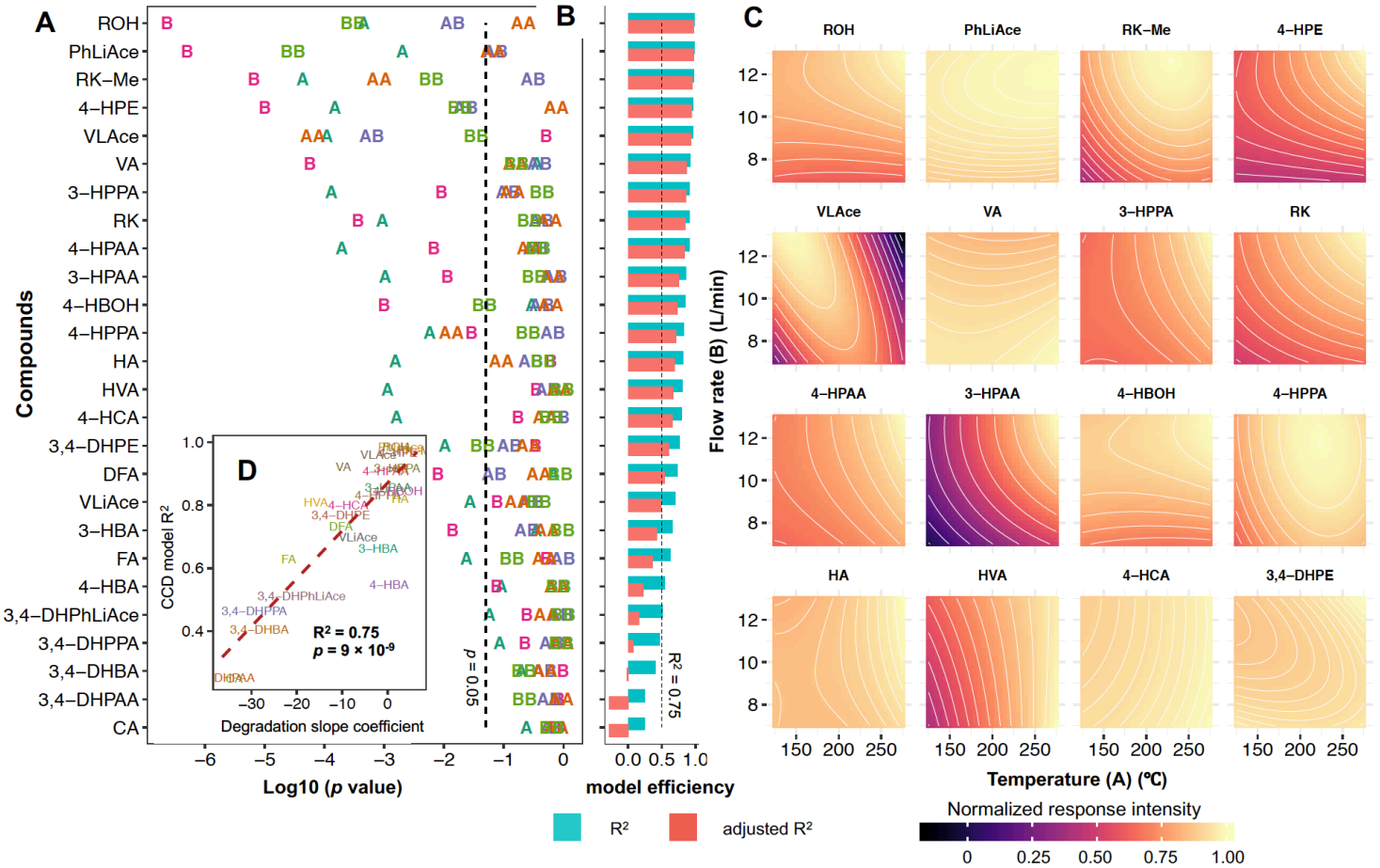


Figure D-2. Optimization of electrospray ionization (ESI) drying gas temperature (DGT, factor A) and drying gas flow rate (DGF, factor B) using central composite design (CCD) quadratic model. Plot (**A**) shows the *p* values of term coefficients. For term annotations, letters A and B refer to the main effects of factors A, and B, respectively; AB, the interaction effects; AA and BB, the corresponding quadratic effects. Plot (**B**) displays the model coefficient of determination (R^2) and adjusted R^2. Plot (**C**) shows representative contour plots. Plot (**D**) presents the correlation of model efficiency with compound stability. Plots (**A**) and (**B**) share the same y-axis. Compounds are arranged in decreasing order of R^2.

3.4. Validation results

All compounds presented ideal limit of quantification down to picograms injected on column, with linear dynamic range spanning across over three orders of magnitude. The majority of compounds including RK had accuracy achieved at 80~120 % in both plasma and brain at four different spiked levels, though for certain compounds accuracy was more inflated or underestimated. Repeatability was mostly below or around 5%, with brain samples presenting more data variability than plasma. Matrix effects, recovery and processing efficiency validated at two spike levels was generally restricted within 80~120%, though brain samples imposed higher challenge to recovery than plasma, and that accuracy-aberrant compounds showed similar drifting behavior with respect to these three validated aspects. Detailed results are shown in **Figure D-3** and **Figure E-4**, and **Table E-5** down to **Table E-9**.

Table D-1. Compounds and dynamic multiple reaction monitoring (dMRM) parameters of the developed method.

Elution order	Compounds	MW (Da)	RT (min)	Pol	Precursor (*m/z*)	Frag (V)	Quantifier (*m/z*)	CE (eV)	Qualifier (*m/z*)	CE (eV)	ratio	Dwell time (ms)	IS
1	4-HBOH	124	1.01	+	107.0	45.0	77.0	24	-	-	-	38	1
2	3,4-DHBA	154	1.02	-	153.0	86.0	109.1	12	108.1	28	25	31	1
3	3,4-DHPE	154	1.06	-	153.1	85.0	123.2	14	122.4	23	6	31	1
4	3,4-DHPAA	168	1.19	-	167.0	60.0	123.1	5	-	-	-	27	1
5	4-HBA	138	1.43	-	137.0	74.0	93.1	16	-	-	-	25	1
6	4-HPE	138	1.42	+	121.1	60.0	77.1	24	51.2	44	50	23	1
7	HA	179	1.56	-	178.1	80.0	134.1	8	77.2	16	37	22	1
8	4-HPAA	152	1.61	-	151.0	70.0	107.2	0	-	-	-	21	1
9	IS-1	142	1.4	-	141.1	73.0	97.1	16	69.1	36	8	32	-
10	3,4-DHPPA	182	1.62	-	181.1	80.0	137.1	9	59.2	13	22	21	1
11	VA	168	1.67	-	167.0	80	152.0	12	108.1	16	30	21	1
12	CA	180	1.69	-	179.0	88.0	135.1	16	89.1	36	2	21	1
13	3-HBA	138	1.8	-	137.0	88.0	93.1	8	-	-	-	21	1
14	HVA	182	1.85	-	181.1	58.0	137.1	4	122.1	12	7	21	1
15	3-HPAA	152	1.86	-	151.0	55.0	107.1	4	65.2	28	6	21	1
16	4-HPPA	166	2.17	-	165.1	78.0	59.1	8	121.1	8	27	22	1
17	4-HCA	164	2.27	-	163.0	80.0	119.1	12	93.1	36	8	23	1
18	DFA	196	2.44	-	195.1	100	136.1	11	121.1	27	39	23	2
19	DHPhLiAce	178	2.48	+	179.1	87.0	143.0	16	115.0	28	52	24	2
20	3-HPPA	166	2.54	-	165.1	88.0	121.1	8	119.1	12	14	24	2
21	FA	194	2.58	-	193.1	88.0	134.1	16	178.1	12	66	25	2
22	ROH	166	2.79	+	107.0	130.0	77.1	24	51.2	40	46	30	2
23	RK	164	3.01	+	107.1	120.0	77.1	21	51.1	37	47	35	2
24	VLAce	194	3.26	+	195.1	60.0	137.1	4	-	-	-	53	2
25	VLiAce	192	3.45	+	193.1	110.0	175.1	12	143.0	16	82	66	2
26	IS-2	155	4.42	-	154.2	74.0	110.1	8	82.2	20	7	108	-
27	PhLiAce	146	4.78	+	147.1	90.0	129.1	12	-	-	-	119	2
28	RK-Me	178	4.86	+	121.0	110.0	78.1	28	77.1	20	64	168	2

For table header abbreviations, MW, molecular weight; RT, retention time; Pol, polarity; Frag, fragmentor voltage; CE, collision energy; IS, internal standard. Compound abbreviations refer to 2.1.The column of "ratio" refers to the expected abundance of qualifier ion relative to the quantifier ion. Each transition was monitored within one-minute time window. The dwell time was calculated based on dMRM mode, depending on the width of detection time window and the extent of retention time overlap.

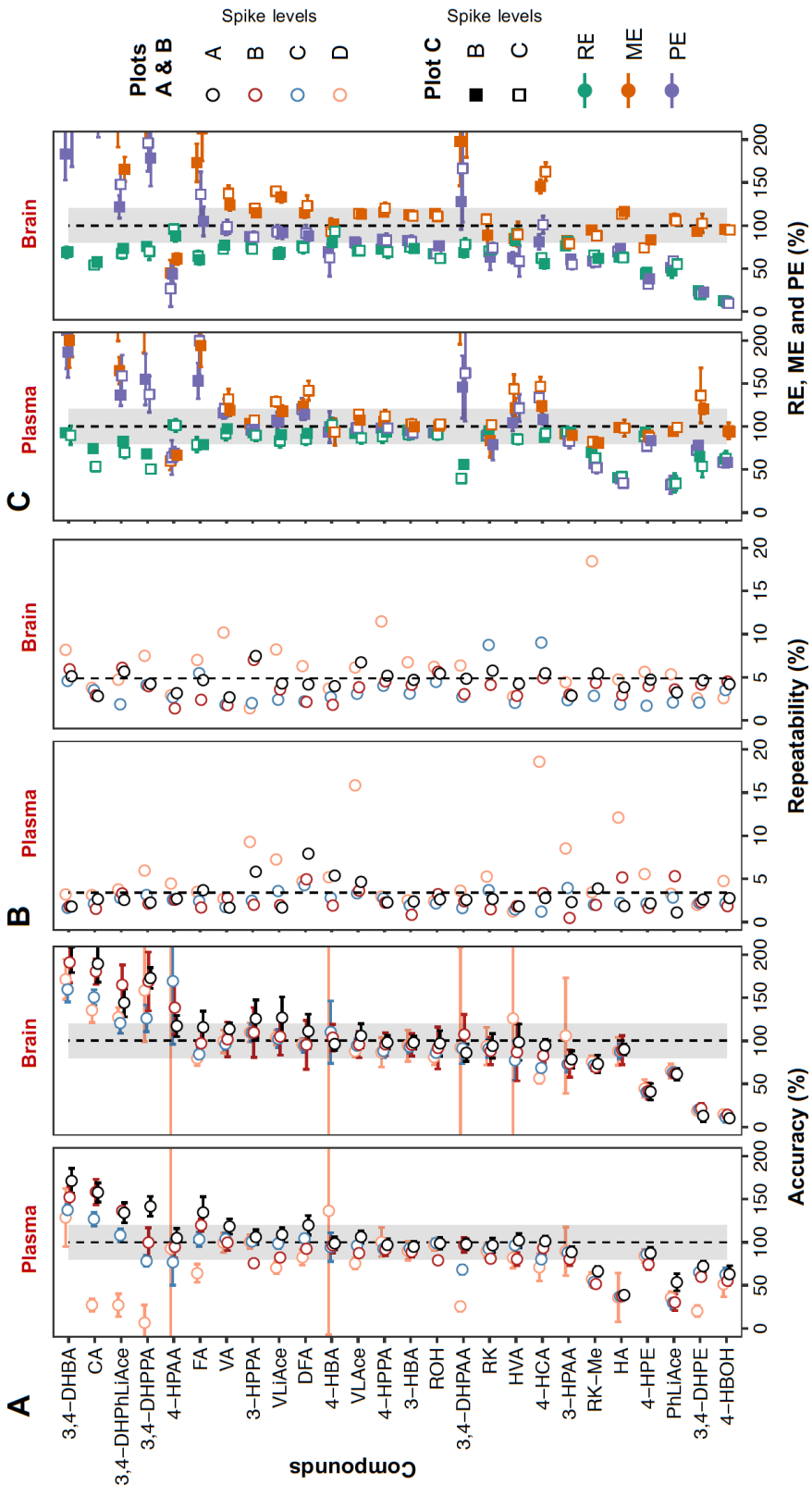


Figure D-3. Method validation results for (**A**), accuracy; (**B**), repeatability; (**C**), recovery (RE), matrix effects (ME) and processing efficiency (PE). Data are expressed as the mean with standard deviation indicated by error bar. The shaded area in plots (**A**) and (**C**) denotes the 80 ~120% region. Compounds are arranged in numerically decreasing order of accuracy. A small number of outliers falling outside of the displayed scale are not shown.

Table D-2. Validated sensitivity, calibration and dynamic linear range.

Elution order	Compound	Sensitivity		Linear range for Low concentration			Linear range for high concentration		
		LOD	LOQ	Range	Calibration curve	R^2	Range	Calibration curve	R^2
1	4-HBOH	0.4	0.8	0.8 ~ 785	y = 1.289 x + 0.006	0.9982	785 ~ 6286	y = 0.959 x + 2.141	0.9951
2	3,4-DHBA	0.7	1.4	1.4 ~ 696	y = 0.548 x + 0.007	0.9982	696 ~ 5571	y = 0.395 x + 0.947	0.9967
3	3,4-DHPE	0.2	0.4	0.4 ~ 720	y = 1.475 x + 0.006	0.9888	720 ~ 5762	y = 0.728 x + 4.320	0.9946
4	3,4-DHPAA	1.4	2.7	2.7 ~ 696	y = 0.607 x + 0.003	0.9989	696 ~ 5571	y = 0.659 x + 0.782	0.9897
5	4-HBA	0.2	0.3	0.3 ~ 678	y = 2.609 x + 0.149	0.9976	678 ~ 5428	y = 2.396 x + 0.411	0.9932
6	4-HPE	0.8	1.6	1.6 ~ 833	y = 0.636 x + 0.002	0.9945	833 ~ 6666	y = 0.384 x + 1.553	0.9959
7	HA	1.3	2.6	2.6 ~ 667	y = 0.510 x + 0.002	0.9984	667 ~ 5333	y = 0.397 x + 0.707	0.9959
8	4-HPAA	1.5	3.1	3.1 ~ 792	y = 0.207 x + 0.002	0.9982	792 ~ 6333	y = 0.104 x + 0.769	0.9923
9	IS-1	-	-	-	-	-	-	-	-
10	3,4-DHPPA	3.0	6.0	6.0 ~ 774	y = 0.394 x + 0.001	0.9955	774 ~ 6190	y = 0.212 x + 1.214	0.9955
11	VA	1.3	2.6	2.6 ~ 673	y = 0.709 x + 0.006	0.9978	673 ~ 5381	y = 0.440 x + 1.598	0.9956
12	CA	1.3	2.6	2.6 ~ 655	y = 0.812 x + 0.004	0.9954	655 ~ 4583	y = 0.843 x + 0.844	0.9868
13	3-HBA	1.5	2.9	2.9 ~ 750	y = 1.064 x + 0.002	0.9997	750 ~ 6000	y = 0.884 x + 1.025	0.9949
14	HVA	2.8	5.6	5.6 ~ 714	y = 0.400 x + 0.005	0.9959	714 ~ 5714	y = 0.193 x + 1.303	0.9944
15	3-HPAA	0.4	0.8	0.8 ~ 815	y = 0.704 x + 0.001	0.9975	815 ~ 6523	y = 0.483 x + 1.415	0.9959
16	4-HPPA	3.0	6.0	6.0 ~ 773	y = 0.260 x + 0.000	0.9994	773 ~ 6190	y = 0.220 x + 0.234	0.9933
17	4-HCA	0.2	0.4	0.4 ~ 732	y = 1.593 x + 0.004	0.9987	732 ~ 5857	y = 1.357 x + 1.147	0.9945
18	DFA	0.8	1.6	1.6 ~ 833	y = 1.287 x + 0.010	0.9980	833 ~ 6666	y = 0.555 x + 6.856	0.9946
19	DHPhLiAce	0.2	0.3	0.3 ~ 351	y = 3.582 x + 0.023	0.9902	351 ~ 5619	y = 2.011 x + 7.074	0.9908
20	3-HPPA	0.7	1.5	1.5 ~ 756	y = 3.363 x + 0.009	0.9971	756 ~ 6047	y = 1.298 x + 17.38	0.9924
21	FA	0.4	0.8	0.8 ~ 786	y = 0.958 x + 0.007	0.9949	786 ~ 6286	y = 0.462 x + 4.093	0.9961
22	ROH	0.3	0.7	0.7 ~ 690	y = 3.723 x + 0.027	0.9952	690 ~ 5523	y = 1.778 x + 14.05	0.9959
23	RK	1.3	2.7	2.7 ~ 684	y = 2.409 x + 0.002	0.9991	684 ~ 5476	y = 1.381 x + 7.716	0.9968
24	VLAce	0.7	1.5	1.5 ~ 759	y = 2.595 x + 0.004	0.9997	759 ~ 6071	y = 1.599 x + 8.717	0.9973
25	VLiAce	0.2	0.4	0.4 ~ 774	y = 4.055 x + 0.019	0.9972	774 ~ 6190	y = 1.837 x + 18.78	0.9952
26	IS-2	-	-	-	-	-	-	-	-

27	PhLiAce	0.7	1.4	1.4 ~ 726	y = 3.372 x + 0.001	0.9992	726 ~ 5809	y = 2.668 x + 5.473	0.9964
28	RK-Me	0.8	1.7	1.7 ~ 851	y = 1.423 x + 0.002	0.9995	851 ~ 6809	y = 0.889 x + 5.042	0.9954

Compound abbreviations refer to section 2.1. Range, limit of detection (LOD) and limit of quantification (LOQ) are in the unit of ng/mL. The injection volume was 3.5 µL.

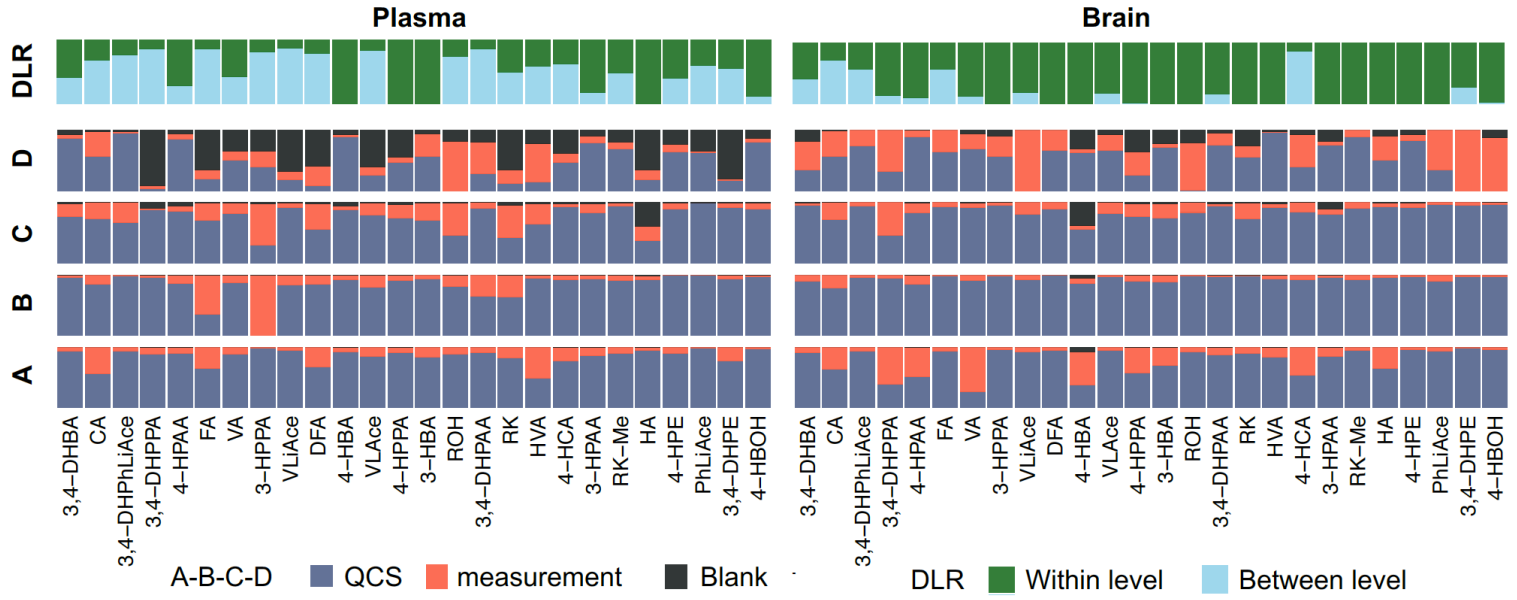


Figure D-4. Accuracy variance partition based on error propagation rule and random effects analysis of variance (RND-ANOVA). The variance is partitioned as percentage into blank sample analysis variability (n' = 3), intrinsic differences in quality control samples (QCS) (a=5) and pure measurement error (n=2). Compounds are arranged in numerically decreasing order of accuracy

3.5. Accuracy inference with RND-ANOVA at spike levels

The following statistical consideration motivated the use of RND-ANOVA for accuracy validation: a reliable validation may be always favored by preparation of many QCS (the factor), ideally by an imaginary pool of QCS's of infinite size (the population), while in practice the QCS prepared essentially represents only a *random sample* (a = 5) drawn from the infinite population, with each QCS being a *random treatment* or *level*. With such experimental limitation in mind, RND-ANOVA was applied to make generalization or inference to the method based on the random QCS prepared. The RND-ANOVA-derived accuracy variation was mostly *ca* 5% higher than otherwise not used (**Figure E-5**).

Another important function of RND-ANOVA is variance partition. The total variance of accuracy, by law of error propagation, could be split and attributed to errors respectively from blank samples (n' = 3) and QCS's; and the latter could be further split by RND-ANOVA and attributed to intrinsic differences in QCS's (a = 5) and pure measurement error (n = 2), as shown in **Figure D-4** (prior page). The variance attributed to QCS's reflected spiking inconsistency and sample inhomogeneity. The variance due to measurement error mirrored within biomatrix instrumental imprecision, integration-associated inconsistency, and compound liability during the 10-hour period between injection repetitions; as injections were made in complete randomized order across different spike concentrations, such measurement error also incorporated carryover effects.

3.6. Accuracy inference with RND-ANOVA throughout calibration range

While accuracy was routinely validated at several representative *discrete* concentration levels, the analyte concentration in an unknown sample in practice could

reside anywhere across the *continuous* scale of calibration beyond the validated levels. As such, RND-ANOVA was applied for accuracy inference across the entire calibration range, with the mean square error associated with within-level variability substituted by the pooled accuracy variance across spike levels. The accuracy inference is shown in **Figure D-4** and **Figure E-6**.

3.7. Repeatability and precision

Repeatability can be conveniently derived from accuracy validation based on the associated mean square error. Compared with "precision" in literature, which is typically conducted with multiple injections on the *same* sample, repeatability takes reduced repetitions ($n = 2$) on each sample and subjects the samples across *different* QCS's ($a = 5$) (variation in QCS's *per se* was then partitioned out) with randomized injection order. Therefore, repeatability validated in this work essentially constitutes precision yet scrutinized in a more complete context, comprehensively reflecting errors from multiple sources (see QCS's measurement error as mentioned above) and better reflects the true consistency in a real batch analysis. In addition, an "anatomized" analysis of repeatability appears to suggest an analyte-specific instrumental drifting over the batch time (**Figure E-7**).

3.8. Analytical method correlation analysis

The analytical and statistical results validated and computed in this work were subjected to a comprehensive pairwise correlation analysis to understand the underlying mechanism of method performance, with an overview presented as a multivariate

correlation matrix heatmap in **Figure D-5A**. Below follows a brief discussion of some key perspectives.

Accuracy was positively correlated with recovery, which reflects analyte extraction efficiency; and positively correlated with matrix effects. This mirrors suppression or enhancement of analyte ionization in the ESI chamber caused by co-extracted biomatrical compounds and as a result, positively correlated with processing efficiency, which is the multiplied product of and therefore a reflection of combined effects of matrix effect and recovery, accounting for 74~84% of accuracy levels in plasma and brain samples (**Figure D-6A**). As processing efficiency was calculated based on peak area, it essentially constitutes the accuracy without use of IS for correction of extraction loss and matrix effects. As such, the difference between accuracy and processing efficiency reflects the correction efficacy of IS. To analyze such efficacy, IS correction index is introduced and defined here as the absolute deviation of accuracy from one hundred percent divided by such deviation of processing efficiency. Thus, the smaller the index number, the higher correction power the IS exerts. The IS correction index was positively correlated with accuracy determined for the brain tissue with about 58% contribution, yet less so in plasma (**Figure D-7**).

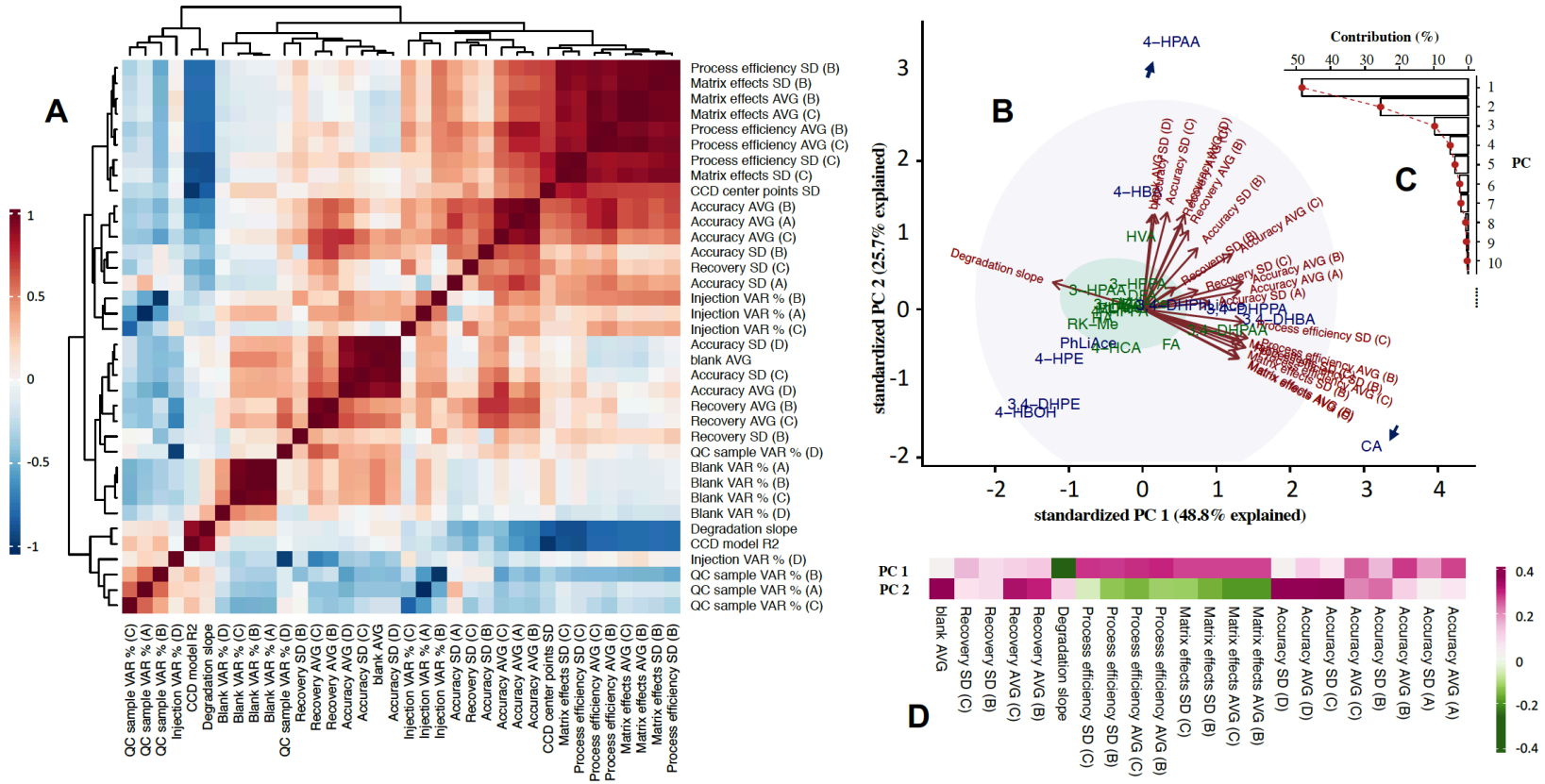


Figure D-5. Multivariate analysis on method analytical and validation results in mice *brain* samples. (**A**), heatmap of correlation matrix of analytical merits; (**B**), principle component analysis (PCA) of all analytes based on selected validation parameters with loadings display; (**C**), contribution percentage of constructed principle components (PCs) to total data variability; (**D**), eigenvector matrix. Analytes with averaged accuracy of 80~120% are marked in green, and outside this range in blue, with ellipse of corresponding color denoting *ca* 70% normal distribution range. For abbreviations applied in the plots, QC, quality control; VAR, variance; CCD, central composite design; SD, standard deviation; AVG, the average level.

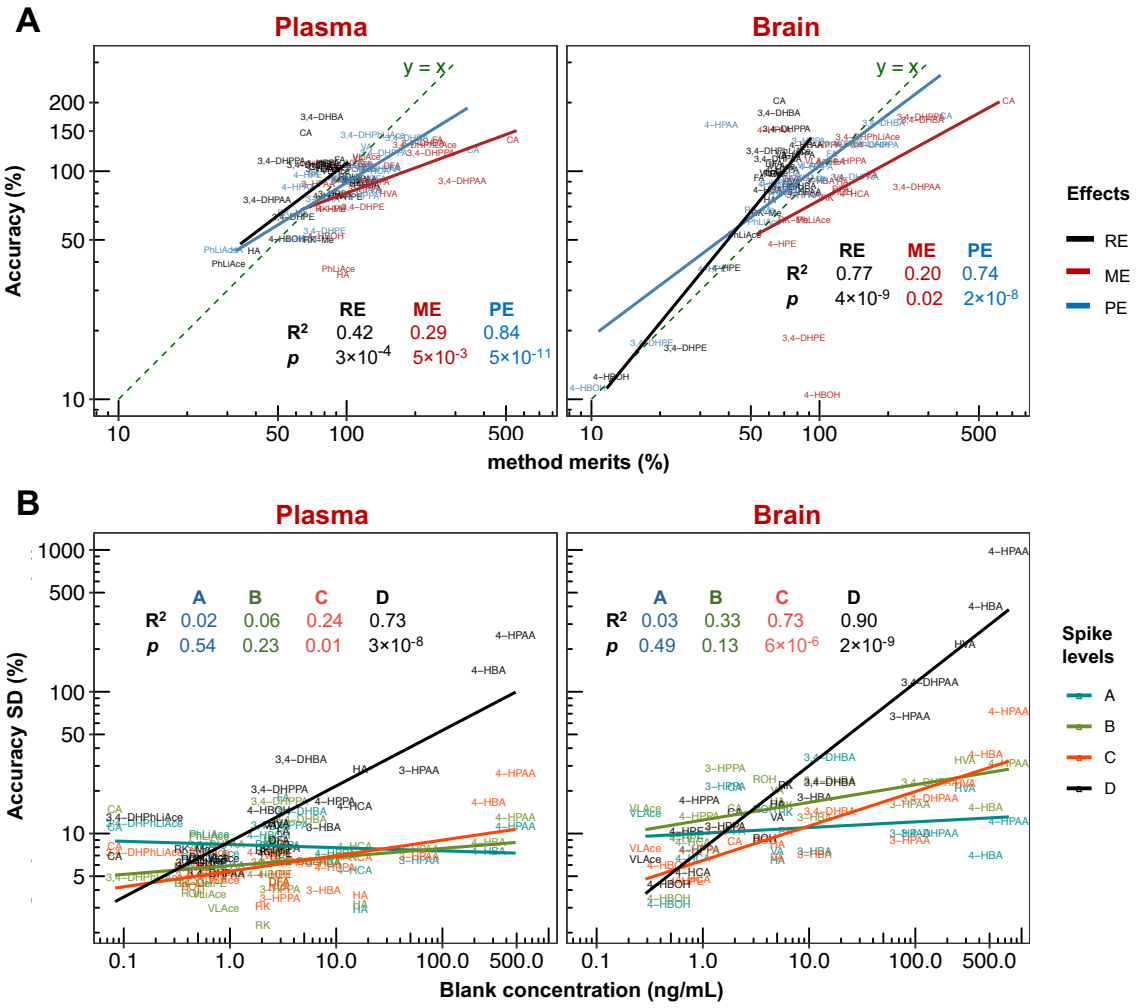


Figure D-6. Correlation analysis of validated accuracy mean value and standard deviation (SD) with recovery (RE), matrix effect (ME) and processing efficiency (PE) (**A**), and with endogenous or background concentration in the biomatrices (**B**). In plot (**A**), accuracy was averaged across three spike levels of A, B and C (D the lowest level was not counted due to high susceptibility to background interference), and RE, ME and PE were respectively averaged across two levels of B and C. In both plots (**A**) and (**B**), linear regression statistics were calculated based on base-10 logarithmically transformed data.

Accuracy variability was increasingly susceptible to blank concentration at lower spike concentrations. At spike level of d (15 ng/mL spiked in processed sample) in plasma and brain matrices, 73% and 90% of increase in the total accuracy variability could be respectively attributed to error from blank concentration deduction (**Figure D-6B**). This effect rendered quantification imprecise at the lower end of calibration range for

compounds with high concentration in the background, such as 4-HPAA, 4-HBA, 3,4-DHPAA, HVA and 3-HPAA (100 ~ 700 ng/mL in sample). Interestingly, most of the background interference was introduced from the commercial β-glucuronidase solution (extracted from limpets or *Patella vulgata*) used for deconjugating phase II metabolites, while fewer compounds were found truly endogenous in the blank biomatrices. For example, 3,4-DHPAA and HVA were found at high levels in blank brain matrices yet lacking in plasma (**Figure E-8**).

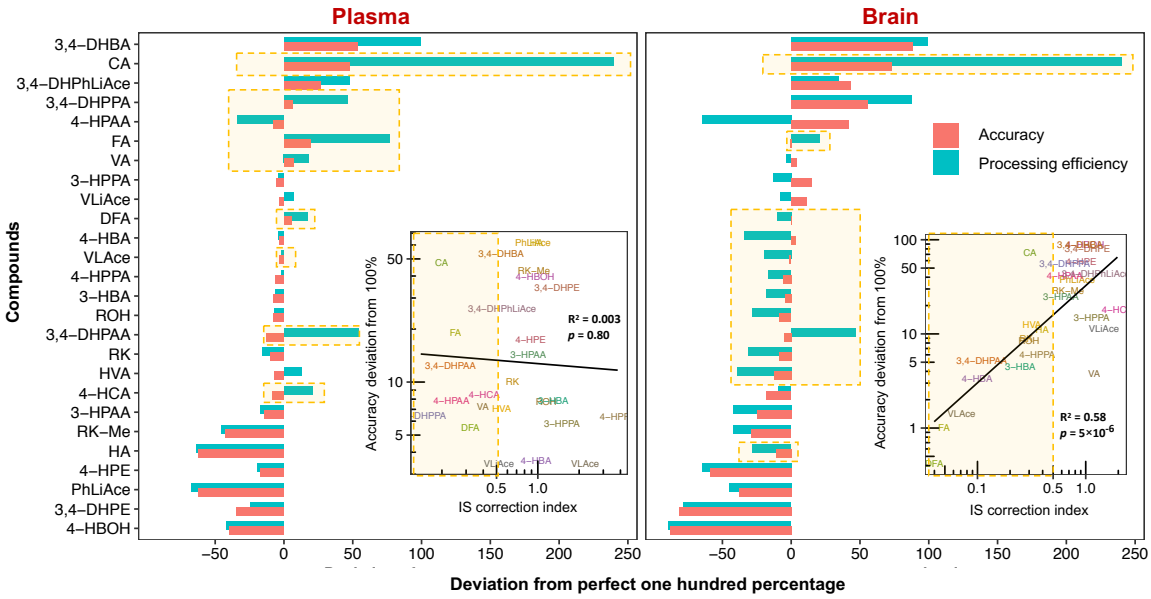


Figure D-7. Analysis of internal standard (IS) correction efficiency. The correction efficiency was manifested by the difference between the accuracy and processing efficiency. Compounds are arranged in numerically decreasing order of accuracy, and compounds with IS correction index smaller than 0.5 are shaded in light orange color. Regression statistics were calculated based on base-10 logarithmically transformed data.

Compound (in)stability is another factor with relevant impact on method development and validation results. Among all analytes investigated, about half featured 2 ~ 36% loss per hour in *pure solvent* at 4 ℃ following a short-term zero-order dynamic model (**Figure D-8A** and **Figure E-3**). In contrast, all compounds presented remarkably elevated stability

in *biomatrices*, manifested by the excellent repeatability which also incorporated compound degradation effects as aforementioned (mostly below 5% error; See **Figure E-7** for additional stability analysis based on repeatability analysis and **Figure E-9** for a separate stability experiment). A similar phenomenon was also noted in some earlier studies [38, 39]. Such (in)stability discrepancy in pure solvent and biomatrices, as a result, increased and contributed to *ca* 65% of the apparent variability of matrix effect and processing efficiency, whose validation involved both pure solvent and biomatrical samples; but with little impact on recovery, which only involved biomatrical samples (**Figure D-8B**). Thus, (in)stability of compounds within a particular biological matrix could also lead to overestimation of the validated accuracy when calibration is prepared in neat solvent (**Figure E-10**, as well as the aforementioned modelling insufficiency in the ESI optimization.

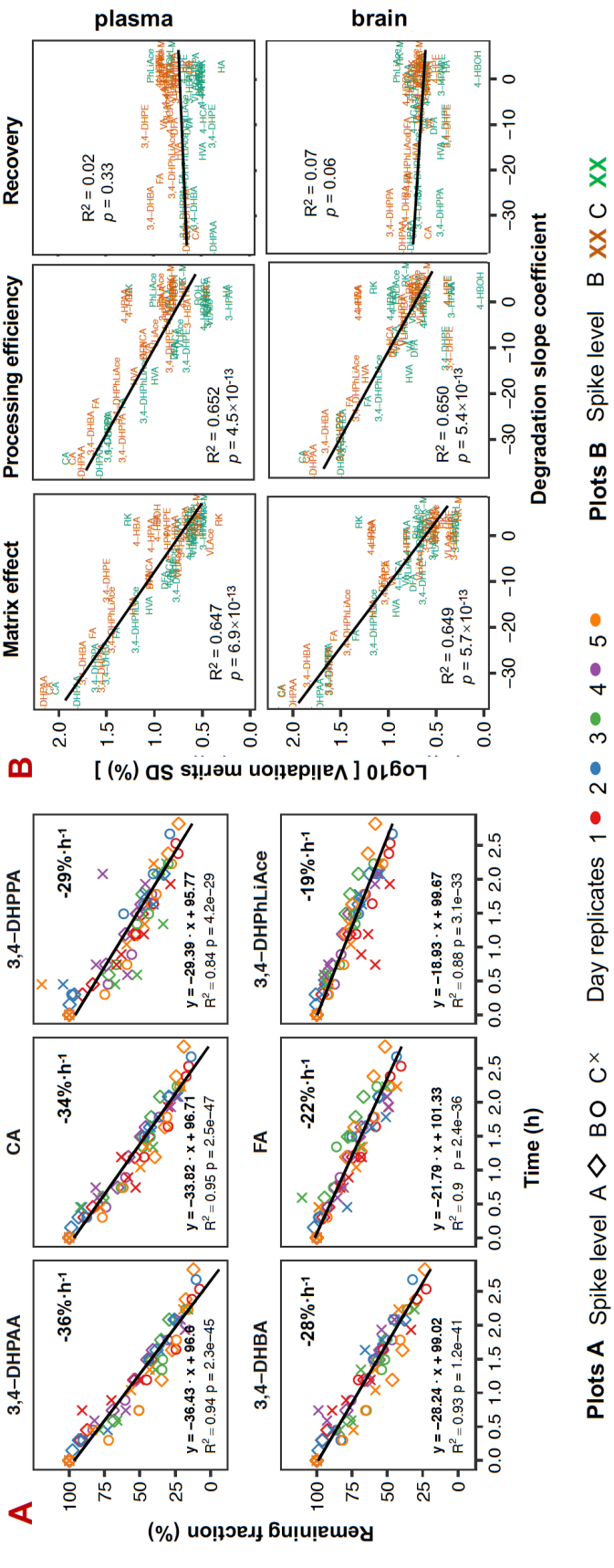


Figure D-8. Compound degradation and its effect on method validation. Plot (**A**) displays the top six most reliable compounds in pure solvent (60% methanol with 0.1% formic acid) stored in 4°C-maintained UHPLC autosampler and zero order dynamic model statistics. Experiments were replicated over five different days and concentrations (A, 2000 ng/mL; B, 1000 ng/mL; and C, 150 ng/mL), and peak areas were normalized to the first injection of each sample of different concentrations for each day as the correspondingly remaining fraction. The slope coefficient corresponds to the compound percent loss per hour. Plot (**B**) shows the correlation of liability (expressed as degradation slope) with matrix effect, processing efficiency and recovery.

3.9. Analytes profile comparison by PCA

All analytes' profile was compared using PCA with selected key validation parameters, as shown in **Figure D-5-B, C** and **D**. Explaining *ca* 74% of total variation, the first two principle components (PCs) reflect the closeness / uniqueness of analytes in context of their analytical characteristics. In contrast, the loading arrows reflect correlations among original variables (OAs, the analytical characteristics), i.e., OAs clustered together are generally positively correlated, those with reverse directions negatively correlated, and those close to perpendicularity only marginally correlated, agreeing with the correlation matrix in **Figure D-5A.** The loadings display the correlation of PCs and OAs and serve as the gateway to PC interpretation. The associated eigenvectors elements (**Figure D-5D**), which are the variance-unadjusted counterpart of loadings, provide a more straightforward approach for PC interpretation, as the eigenvector elements *per se* are the exact coefficients of the linear transformation for PCs' construction, and directly measures the weight and functionality of each OA (in presence of other OAs; while loadings measure the weight ignoring all other OAs by standardizing off the associated variance) in this procedure. As such, the PCs in this work was interpreted mostly using the eigenvectors as the primary tool as discussed below.

PC1 is first and foremost associated with compounds' neat-solvent liability, expressed as the zero-order kinetic degradation slope, with a negative sign (**Figure D-5D)**. As such liability is causally associated with validation performance in terms of accuracy, matrix effects and processing efficiency as aforementioned, PC1 is therefore also associated with these validation results yet with a positive sign, but not as much association with recovery as expected. As such, PC1 essentially constituted the "degradation dimension". This leads

to CA as well as 3, 4-DHBA and 3, 4-DHPPA with their high liability sliding to the right side along the direction of PC1, while 4-HBOH and 3, 4-DHPE with their somehow positive slope slightly shifted to the left side.

PC2 is first of all significantly associated with the background level in the biomatrices. As the background interference propagates to accuracy determination especially at low levels as discussed above. PC2 is therefore also positively associated with the validated accuracy. As such, PC2 may be interpreted as the "background interference" dimension. Following this, 4-HPAA, due to its high background occurrence in enzyme solution, was found at the very periphery of the PCA plot along the PC2 direction.

4. Conclusion

A UHPLC-QqQ-MS/MS method for RK and 25 analytes identified as RK-derived metabolites was developed and validated. Design of experiment methodology was applied for efficient method optimization. Application of RND-ANOVA, a universal correlation analysis and PCA diagnosis revealed how the multiple parameters contributed to method performance. Two particular phenomena, the analytes' background occurrence in the commercial enzyme solution used for metabolites deconjugation, and the unexpected rapid degradation of analytes under 4 °C in pure solvent vs. elevated stability in biomatrices, constituted the essence of the first two PCA dimensions, exerting crucial impact on method performance. In view of the validation results, the proposed method could serve for studies on RK metabolism, pharmacokinetics and bioavailability and associated safety / toxicity evaluation using *in vivo* models or in clinical trials.

5. References

[1] Aprea E, Biasioli F, Gasperi F. Volatile compounds of raspberry fruit: from analytical methods to biological role and sensory impact. Molecules 2015;20:2445-74.

[2] Larsen M, Poll L, Callesen O, Lewis M. Relations between the content of aroma compounds and the sensory evaluation of 10 raspberry varieties (*Rubus idaeus* L). Acta Agric Scand 1991;41:447-54.

[3] Honkanen E, Pyysalo T, Hirvi T. The aroma of finnish wild raspberries, Rubus idaeus, L. Zeitschrift für Lebensmitteluntersuchung und-Forschung A 1980;171:180-182.

[4] Bredsdorff L, Wedebye EB, Nikolov NG, Hallas-Moller T, Pilegaard K. Raspberry ketone in food supplements--High intake, few toxicity data--A cause for safety concern? Regul Toxicol Pharmacol 2015;73:196-200.

[5] Beekwilder J, van der Meer IM, Sibbesen O, Broekgaarden M, Qvist I, Mikkelsen JD, Hall RD. Microbial production of natural raspberry ketone. Biotechnol J 2007;2:1270-9.

[6] Wang L, Meng X, Zhang F. Raspberry ketone protects rats fed high-fat diets against nonalcoholic steatohepatitis. J Med Food 2012;15:495-503.

[7] Morimoto C, Satoh Y, Hara M, Inoue S, Tsujita T, Okuda H. Anti-obese action of raspberry ketone. Life Sci 2005;77:194-204.

[8] Cotten BM, Diamond SA, Banh T, Hsiao YH, Cole RM, Li J, Simons CT, Bruno RS, Belury MA, Vodovotz Y. Raspberry ketone fails to reduce adiposity beyond decreasing food intake in C57BL/6 mice fed a high-fat diet. Food Funct 2017;8:1512-1518.

[9] Park KS. Raspberry ketone, a naturally occurring phenolic compound, inhibits adipogenic and lipogenic gene expression in 3T3-L1 adipocytes. Pharm Biol 2015;53:870-5.

[10] Park KS. Raspberry ketone increases both lipolysis and fatty acid oxidation in 3T3-L1 adipocytes. Planta Med 2010;76:1654-1658.

[11] Vrhovsek U, Masuero D, Gasperotti M, Franceschi P, Caputi L, Viola R, Mattivi F. A versatile targeted metabolomics method for the rapid quantification of multiple classes of phenolics in fruits and beverages. J Agric Food Chem 2012;60:8831-40.

[12] Zhao D, Bo Yuan; E. Carry; G.M. Pasinetti; L. Ho; J. Faith; I. Mogno; J.E. Simon and Q.L.Wu. . Development and validation of an ultra-high performance liquid chromatography/triple quadrupole tandem mass spectrometry method for analyzing microbial-derived grape polyphenol metabolites. Manuscript in revision. . J Chromatogr B 2018.

[13] Johnson AR, Carlson EE. Collision-Induced Dissociation Mass Spectrometry: A Powerful Tool for Natural Product Structure Elucidation. Anal Chem 2015;87:10668-78.

[14] Martendal E, Budziak D, Carasek E. Application of fractional factorial experimental and Box-Behnken designs for optimization of single-drop microextraction of 2,4,6-

trichloroanisole and 2,4,6-tribromoanisole from wine samples. J Chromatogr A 2007;1148:131-6.

[15] ChemCalc: a building block for tomorrow's chemical infrastructure. Patiny, Luc; Borel, Alain Journal of Chemical Information and Modeling 2013. DOI: 10.1021/ci300563h.

[16] Fredenhagen A, Derrien C, Gassmann E. An MS/MS library on an ion-trap instrument for efficient dereplication of natural products. Different fragmentation patterns for [M + H]+ and [M + Na]+ ions. J Nat Prod 2005;68:385-91.

[17] Borejsza-Wysocki W, Goers SK, McArdle RN, Hrazdina G. (p-Hydroxyphenyl) butan-2-one levels in raspberries determined by chromatographic and organoleptic methods. Journal of agricultural and food chemistry 1992;40:1176-1177.

[18] Maquin F, Meili M, Chaveron H. Determination of 4-(p-hydroxyphenyl)-2-butanone by mass fragmentometry. Ann Falsif Expert Chim Toxicol 1981;74:511-21.

[19] Leu SY, Chen YC, Tsai YC, Hung YW, Hsu CH, Lee YM, Cheng PY. Raspberry Ketone Reduced Lipid Accumulation in 3T3-L1 Cells and Ovariectomy-Induced Obesity in Wistar Rats by Regulating Autophagy Mechanisms. J Agric Food Chem 2017;65:10907-10914.

[20] Ogawa Y, Akamatsu M, Hotta Y, Hosoda A, Tamura H. Effect of essential oils, such as raspberry ketone and its derivatives, on antiandrogenic activity based on in vitro reporter gene assay. Bioorg Med Chem Lett 2010;20:2111-2114.

[21] Lin CH, Ding HY, Kuo SY, Chin LW, Wu JY, Chang TS. Evaluation of in vitro and in vivo depigmenting activity of raspberry ketone from Rheum officinale. Int J Mol Sci 2011;12:4819-35.

[22] Jeong JB, Jeong HJ. Rheosmin, a naturally occurring phenolic compound inhibits LPS-induced iNOS and COX-2 expression in RAW264.7 cells by blocking NF-kappaB activation pathway. Food Chem Toxicol 2010;48:2148-53.

[23] Khan V, Sharma S, Bhandari U, Sharma N, Rishi V, Haque SE. Suppression of isoproterenol-induced cardiotoxicity in rats by raspberry ketone via activation of peroxisome proliferator activated receptor-alpha. Eur J Pharmacol 2019;842:157-166.

[24] Lee J. Further research on the biological activities and the safety of raspberry ketone is needed. NFS Journal 2016;2:15-18.

[25] Sporstøl S, Scheline R. The metabolism of 4-(4-hydroxyphenyl) butan-2-one (raspberry ketone) in rats, guinea-pigs and rabbits. Xenobiotica 1982;12:249-257.

[26] Zhao D, Yuan B, Kshatriya D, Bello NT, Simon JE, Wu Q. Influence of diet-induced obesity on the bioavailability and metabolism of raspberry ketone (4-(4-hydroxyphenyl)-2-butanone) in mice. Molecular Nutrition and Food Research 2020.

[27] Yuan B, Zhao D, Du R, Kshatriya D, Bello NT, Simon JE, Wu Q. A highly sensitive ultra-high performance liquid chromatography/tandem mass spectrometry method with in-source fragmentation for rapid quantification of raspberry ketone. J Food Drug Anal 2018.

[28] Montgomery DC. Design and analysis of experiments: John wiley & sons; 2017.

[29] Zhao D, Yuan B, Carry E, Pasinetti GM, Ho L, Faith J, Mogno I, Simon J, Wu Q. Development and validation of an ultra-high performance liquid chromatography/triple quadrupole mass spectrometry method for analyzing microbial-derived grape polyphenol metabolites. J Chromatogr B Analyt Technol Biomed Life Sci 2018;1099:34-45.

[30] Food-and-Drug-Administration U. Guidance for Industry: Bioanalytical Method Validation. 2018.

[31] Kruve A, Rebane R, Kipper K, Oldekop ML, Evard H, Herodes K, Ravio P, Leito I. Tutorial review on validation of liquid chromatography-mass spectrometry methods: part II. Anal Chim Acta 2015;870:8-28.

[32] Matuszewski BK, Constanzer ML, Chavez-Eng CM. Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC-MS/MS. Anal Chem 2003;75:3019-30.

[33] Taylor PJ. Matrix effects: the Achilles heel of quantitative high-performance liquid chromatography-electrospray-tandem mass spectrometry. Clin Biochem 2005;38:328-34.

[34] Johnson RA, Wichern DW. Applied multivariate statistical analysis: Prentice hall Upper Saddle River, NJ; 2002.

[35] R-Core-Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/. 2013.

[36] Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 2016;32:2847-9.

[37] Ostrowski W, Wojakowska A, Grajzer M, Stobiecki M. Mass spectrometric behavior of phenolic acids standards and their analysis in the plant samples with LC/ESI/MS system. J Chromatogr B Analyt Technol Biomed Life Sci 2014;967:21-7.

[38] Wang YJ, Pan MH, Cheng AL, Lin LI, Ho YS, Hsieh CY, Lin JK. Stability of curcumin in buffer solutions and characterization of its degradation products. J Pharm Biomed Anal 1997;15:1867-76.

[39] Krook MA, Hagerman AE. Stability of Polyphenols Epigallocatechin Gallate and Pentagalloyl Glucose in a Simulated Digestive System. Food Res Int 2012;49:112-116.

APPENDIX E. SUPPORTING MATERIAL OF APPENDIX D

Supporting material

UHPLC-QqQ-MS/MS method development and validation with statistical analysis: determination of raspberry ketone metabolites in mice plasma and brain

Validation calculations

Accuracy at a given spike level was calculated as:

$$AC\ (\%) = (MC - BC) \times 100\ /\ SC \quad (1)$$

$$S^2(AC)(\%^2) = \left(S^2(MC) + S^2(BC)\right) \times 100^2\ /\ SC^2 \quad (2)$$

where *MC* and *SC* respectively refer to the measured and spiked concentration in quality control samples (QCS's), and *BC* is the background concentration measured in blank matrices. S^2 denotes the variance, and *S* the standard deviation. $S^2(MC)$ was calculated using random effects analysis of variance (RND-ANOVA) as below:

$$MS_{inj} = \sum_{i=1}^{a}\sum_{j=1}^{n}\left(MC_{ij} - \overline{MC}_{i\cdot}\right)^2 / (an - a) \quad (3)$$

$$MS_{QC} = \sum_{i=1}^{a} n\left(\overline{MC}_{i\cdot} - \overline{MC}_{\cdot\cdot}\right)^2 / (a - 1) \quad (4)$$

$$S^2(inj) = MS_{inj}, S^2(QC) = (MS_{QC} - \mathrm{MS}_{inj}) / n \quad (5)$$

$$S^2(MC) = S^2(inj) + S^2(QC) \quad (6)$$

where MS_{inj} and MS_{QC} respectively refer to the squared mean associated with injection repetitions (n = 2) and quality control sample (QCS) replicates (a = 5); MC_{ij} refers to the calculated concentration of the j^{th} injection (j = 1, 2…n) of the i^{th} sample (i = 1, 2…a); $\overline{MC}_{i\cdot}$, the calculated concentration of the i^{th} sample averaged across injection repetitions; and $\overline{MC}_{\cdot\cdot}$, the grand mean of concentration averaged across all QCS and injection repetitions.

Repeatability was calculated as:

$$RP(\%) = S(inj) \times 100\ /\ \overline{MC}_{\cdot\cdot} \quad (7)$$

Matrix effects, recovery and processing efficiency was calculated as below:

$$RE\ (\%) = PrB \times 100\ /\ PoB \quad (8)$$

$$\mathrm{S}^2(RE)(\%^2) = RE^2 \times \left(\frac{S^2(PrB)}{PrB^2} + \frac{S^2(PoB)}{PoB^2}\right) \quad (9)$$

$$ME\ (\%) = (PoB - BK) \times 100\ /\ NS \quad (10)$$

$$\mathrm{S}^2(ME)(\%^2) = ME^2 \times \left(\frac{S^2(PoB) + S^2(BK)}{(PoB - BK)^2} + \frac{S^2(NS)}{NS^2}\right) \quad (11)$$

$$PE\ (\%) = (PrB - BK) \times 100\ /\ NS \quad (12)$$

$$S^2(PE)(\%^2) = PE^2 \times \left(\frac{S^2(PrB) + S^2(BK)}{(PrB - BK)^2} + \frac{S^2(NS)}{NS^2}\right) \quad (13)$$

where PrB refers to the peak area of analytes spiked *pre-extraction* in blank biomatrices (being the same set of QCS's used for accuracy validation); PoB, the peak area of analytes spiked *post-extraction* in blank

biomatrices; NS, the peak area of analytes spiked in pure solvent (60 % methanol with 0.1% formic acid); BK, the peak area in blank biomatrices (being the same blank samples used in accuracy validation).

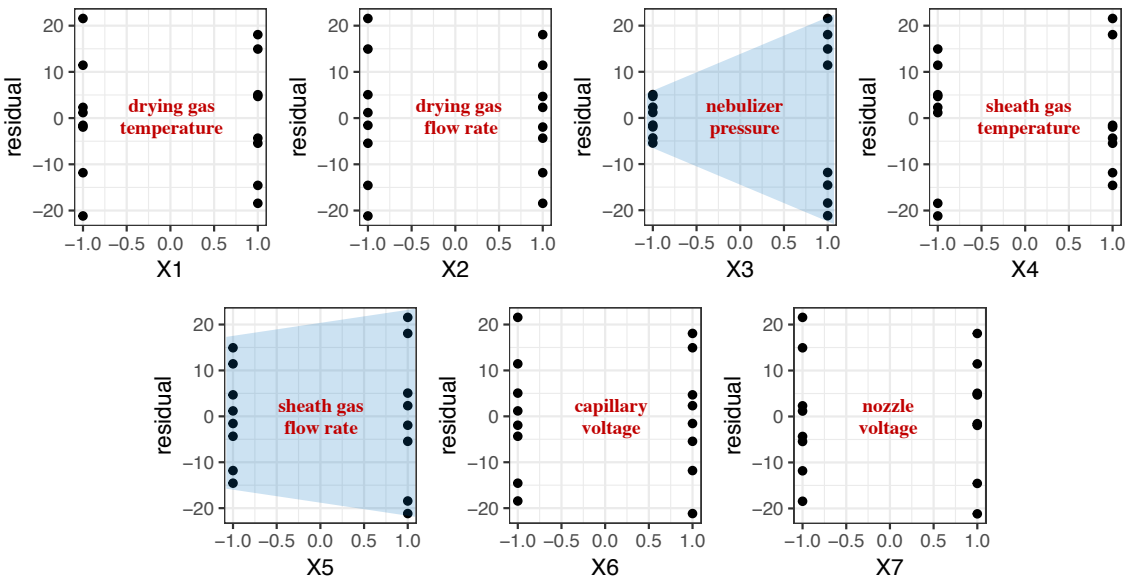


Figure E-1. Dispersion analysis of prior RK-oriented ESI 2_{IV}^{7-3} fractional factorial design. Residuals were calculated based on the reported model. Notice the dispersion effect at higher levels of nebulizer pressure and of, to a lesser extent, sheath gas flow rate, as highlighted by the shaded trapezoid. The assumed residual normality was examined by Shapiro-Wilk test (W = 0.974, p = 0.9014) and quantile-quantile plot and considered satisfied. For experimental levels (-1 / +1) in the prior work, X_1, 250 / 350 °C; X_2, 8 / 13 L/min; X_3, 25 / 40 psi; X_4, 250 / 350 °C; X_5, 8 / 12 L/min; X_6, 1500 / 3000 V; X_7, 500 / 1500 V.

Table E-1. Projection of original 2_{IV}^{7-3} fractional factorial design into 2^3 full factorial design with two replicates

T.C.	replicate	projected "new" design A* X_1	B* X_2	C* X_7	response	predicted**	residual**	original DOE condition Standard order	removed factors (merged in residuals) X_3	X_4	X_5	X_6
1	1	-1	-1	-1	216	219.44	-3.4	13	1	1	1	-1
	2	-1	-1	-1	211	219.44	-8.4	1	-1	-1	-1	-1
2	1	1	-1	-1	295	260.44	34.6	6	1	-1	-1	1
	2	1	-1	-1	270	260.44	9.6	10	-1	1	1	1
3	1	-1	1	-1	371	339.44	31.6	3	-1	-1	1	1
	2	-1	1	-1	333	339.44	-6.4	15	1	1	-1	1
4	1	1	1	-1	208	236.19	-28.2	12	-1	1	-1	-1
	2	1	1	-1	207	236.19	-29.2	8	1	-1	1	-1
5	1	-1	-1	1	237	305.06	-68.1	5	1	-1	1	1
	2	-1	-1	1	385	305.06	79.9	9	-1	1	-1	1
6	1	1	-1	1	373	346.06	26.9	14	1	1	-1	-1
	2	1	-1	1	275	346.06	-71.1	2	-1	-1	1	-1
7	1	-1	1	1	414	425.06	-11.1	7	1	-1	-1	-1
	2	-1	1	1	411	425.06	-14.1	11	-1	1	1	-1
8	1	1	1	1	357	321.81	35.2	16	1	1	1	1
	2	1	1	1	344	321.81	22.2	4	-1	-1	-1	1

T.C., factorial treatment combination. *, letter notations. **, calculation was based on the reduced model (see **Table E-3**). For variable notation, X_1 for drying gas temperature, X_2 drying gas flow rate, X_3 nebulizer pressure, X_4 sheath gas temperature, X_5 sheath gas flow rate, X_6 capillary voltage, and X_7 nozzle voltage.

Table E-2. ANOVA table of *full* model of projected design

Source	Model								Residual	Total
	A	B	C	AB	AC	BC	ABC	subtotal		
SS	3875.06	9168.06	29326.56	20808.06	175.56	1040.06	4795.56	69188.94	16890.5	86079.44
df	1	1	1	1	1	1	1	7	8	15
MS	3875.06	9168.06	29326.56	20808.06	175.56	1040.06	4795.56	9884.13	2111.31	5738.63
F	1.84	4.34	13.89	9.86	0.08	0.49	2.27	4.68	-	-
Prob > F	0.213	0.071	0.006	0.014	0.78	0.503	0.17	0.023	-	-
Contribution	4.50%	10.70%	34.10%	24.20%	0.20%	1.20%	5.60%	80.40%	19.60%	100.00%

Note: A, drying gas temperature; B, drying gas flow rate; C, nozzle voltage.

Table E-3. ANOVA table of *selected* model of projected design

Source	Model					Residual			Total
	A	B	C	AB	subtotal	lack of fit	pure error	subtotal	
SS	3875.06	9168.06	29326.56	20808.06	63177.75	6011.19	16890.5	22901.69	86079.44
df	1	1	1	1	4	3	8	11	15
MS	3875.06	9168.06	29326.56	20808.06	15794.44	2003.73	2111.31	2081.97	5738.63
F	1.86	4.4	14.09	9.99	7.59	0.95	-	-	-
Prob > F	0.2	0.06	0.003	0.009	0.003	0.462	-	-	-
Contribution	4.50%	10.70%	34.10%	24.20%	73.40%	7.00%	19.60%	26.60%	100.00%

Note: A, drying gas temperature; B, drying gas flow rate; C, nozzle voltage.

Table E-4. Central composite design (CCD) of drying gas temperature (DGT, factor A) and flow rate (DGF, factor B).

Standard order	run order	Design points	Coded levels		Actual levels	
			A	B	A (°C)	B (L/min)
1	7	Factorial point	-1	-1	200	8
2	6		1	-1	300	8
3	5		-1	1	200	12
4	9		1	1	300	12
5	11	Star/axial points	-1.414	0	179	10
6	4		1.414	0	320	10
7	8		0	-1.414	250	7.2
8	2		0	1.414	250	12.8
9	12	Center points	0	0	250	10
10	1		0	0	250	10
11	10		0	0	250	10
12	3		0	0	250	10
13	13		0	0	250	10

The other electrospray ionization (ESI) settings were: nebulizer pressure 30 psi; sheath gas temperature 250 °C, with its flow rate 8 L/min; capillary voltage + 3000 V / -2500 V and nozzle voltage + 1500 V / -1000V.

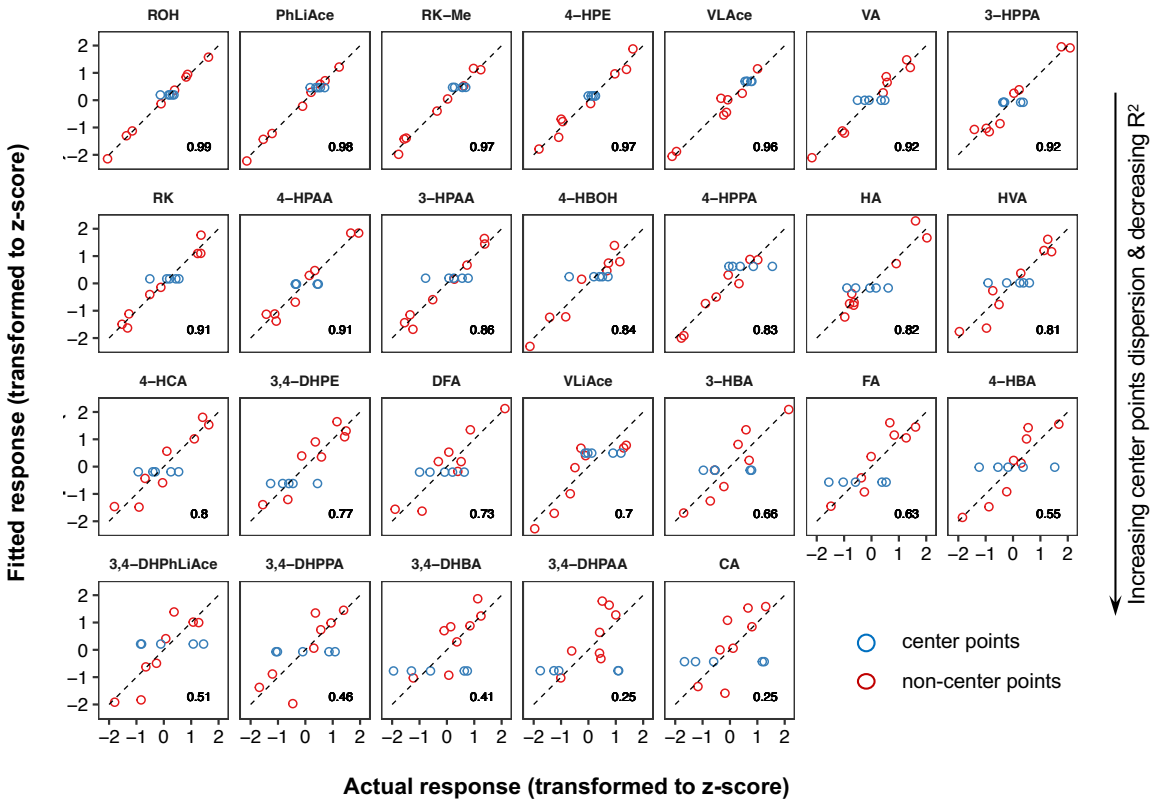


Figure E-2. Scatterplot of fitted vs. actual response of analytes by central composite design (CCD) quadratic model in electrospray ionization (ESI) optimization. Model coefficient of determination (R^2) is noted at the bottom right corner of each faceted plot. Compounds are arranged in decreasing order of R^2. Notice the increasing dispersibility of center points (repeated measurements under the same instrumental settings) accompanying the decrease in R^2. Further analysis shows compound degradation (**Figure E-3**) over the batch time as the major cause of large dispersion of center points.

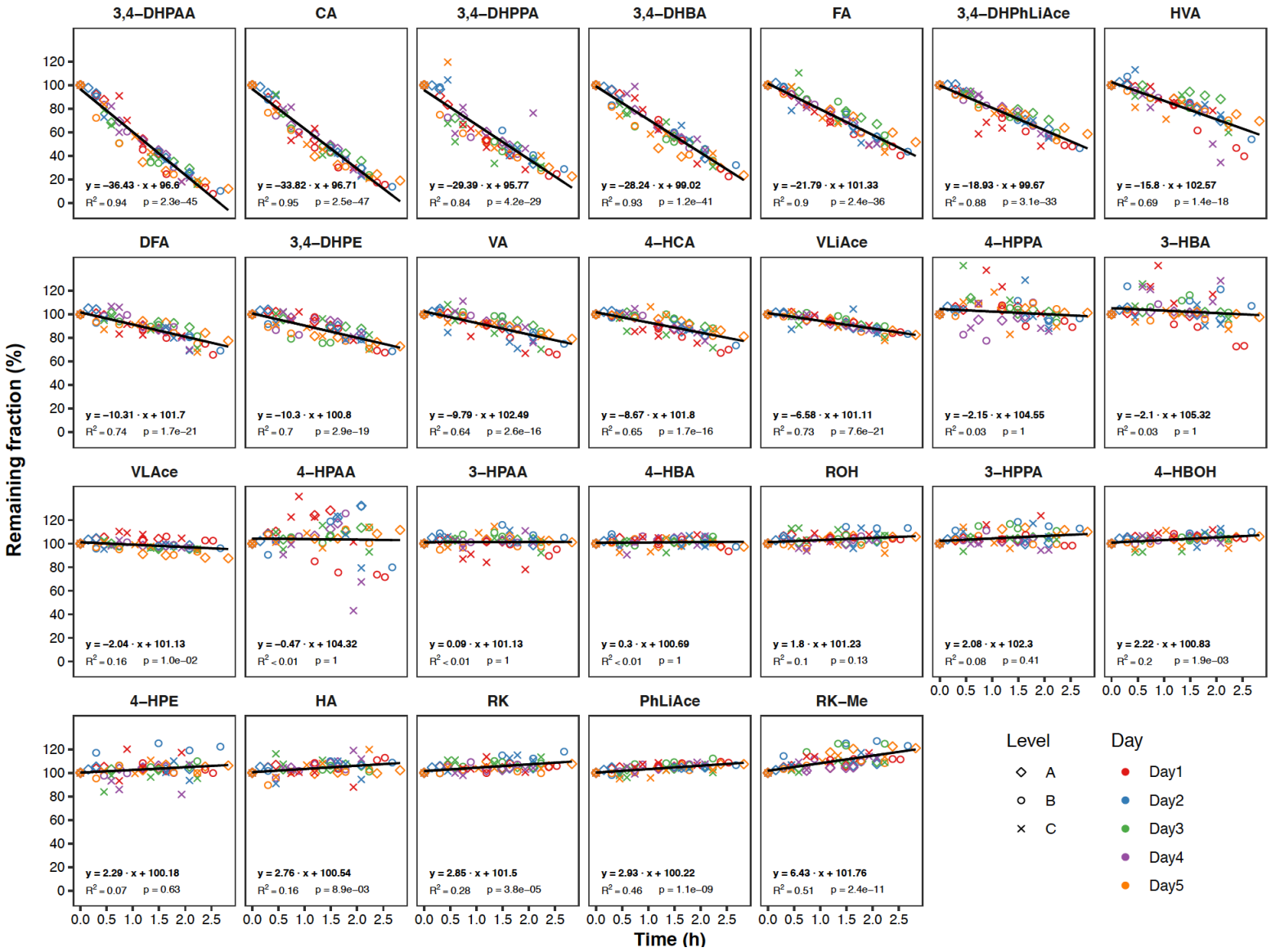

3,4–DHPAA
y = −36.43 · x + 96.6
R² = 0.94 p = 2.3e−45
CA
y = −33.82 · x + 96.71
R² = 0.95 p = 2.5e−47
3,4–DHPPA
y = −29.39 · x + 95.77
R² = 0.84 p = 4.2e−29
3,4–DHBA
y = −28.24 · x + 99.02
R² = 0.93 p = 1.2e−41
FA
y = −21.79 · x + 101.33
R² = 0.9 p = 2.4e−36
3,4–DHPhLiAce
y = −18.93 · x + 99.67
R² = 0.88 p = 3.1e−33
HVA
y = −15.8 · x + 102.57
R² = 0.69 p = 1.4e−18
DFA
y = −10.31 · x + 101.7
R² = 0.74 p = 1.7e−21
3,4–DHPE
y = −10.3 · x + 100.8
R² = 0.7 p = 2.9e−19
VA
y = −9.79 · x + 102.49
R² = 0.64 p = 2.6e−16
4–HCA
y = −8.67 · x + 101.8
R² = 0.65 p = 1.7e−16
VLiAce
y = −6.58 · x + 101.11
R² = 0.73 p = 7.6e−21
4–HPPA
y = −2.15 · x + 104.55
R² = 0.03 p = 1
3–HBA
y = −2.1 · x + 105.32
R² = 0.03 p = 1
VLAce
y = −2.04 · x + 101.13
R² = 0.16 p = 1.0e−02
4–HPAA
y = −0.47 · x + 104.32
R² < 0.01 p = 1
3–HPAA
y = 0.09 · x + 101.13
R² < 0.01 p = 1
4–HBA
y = 0.3 · x + 100.69
R² < 0.01 p = 1
ROH
y = 1.8 · x + 101.23
R² = 0.1 p = 0.13
3–HPPA
y = 2.08 · x + 102.3
R² = 0.08 p = 0.41
4–HBOH
y = 2.22 · x + 100.83
R² = 0.2 p = 1.9e−03
4–HPE
y = 2.29 · x + 100.18
R² = 0.07 p = 0.63
HA
y = 2.76 · x + 100.54
R² = 0.16 p = 8.9e−03
RK
y = 2.85 · x + 101.5
R² = 0.28 p = 3.8e−05
PhLiAce
y = 2.93 · x + 100.22
R² = 0.46 p = 1.1e−09
RK–Me
y = 6.43 · x + 101.76
R² = 0.51 p = 2.4e−11
Remaining fraction (%)
Time (h)
Level
A
B
C
Day
Day1
Day2
Day3
Day4
Day5

Figure E-3. Compound degradation profile over 3 hours in pure solvent (60% methanol with 0.1% formic acid) in 4°C- autosampler and zero-order kinetics model statistics. Compounds are displayed in decreasing order of degradation rate. Experiments were replicated over different days and concentrations (A, 2000 ng/mL; B, 1000 ng/mL; and C, 150 ng/mL), and peak areas were normalized to the first injection of each sample and of each day as correspondingly the remaining fraction. The slope of linear regression noted at the bottom of each faceted plot indicates the degradation rate per hour. The positive slopes (2.22 ~ 6.43, $p < 0.05$) might be a result of slight instrumental drifting (see Figure E-7 for additional analysis).

Table E-5. Accuracy (%) validated for 26 analytes in mice plasma and brain at four spike levels A-D.

No.	compound	Plasma				Brain			
		A	B	C	D	A	B	C	D
1	RK	96.4 ± 8.5	80.9 ± 2.3	92.3 ± 3.1	90 ± 7.7	94.1 ± 14.1	88.2 ± 15.8	91.1 ± 10.3	93.8 ± 21.9
2	RK-Me	66.4 ± 5.8	51.6 ± 4.5	53.9 ± 5.9	57.2 ± 7.8	73.2 ± 10	69.2 ± 3.3	71.6 ± 6.3	75.6 ± 6.6
3	PhLiAce	53.6 ± 9.9	30.2 ± 9.6	29.9 ± 7.4	35.5 ± 6.9	61.4 ± 6.6	61.9 ± 3.4	63.2 ± 6	65.3 ± 8.2
4	3,4-DHPhLiAce	134.2 ± 11.8	136.4 ± 4.9	108.1 ± 7.4	27.1 ± 13.1	144.1 ± 16.3	164.9 ± 22.9	120.6 ± 12	127.3 ± 11.1
5	VLAce	106.2 ± 7.3	87.4 ± 3	96.1 ± 4.7	75.4 ± 6.7	105.8 ± 13.9	95.5 ± 15.2	94.4 ± 7.9	88 ± 6.6
6	VLiAce	108.9 ± 8.3	82.3 ± 3.7	98.4 ± 5.8	70.3 ± 6.5	126.9 ± 24.1	104.9 ± 21.5	102.3 ± 10.7	103.9 ± 8.5
7	ROH	98.6 ± 6.8	79 ± 3.8	99 ± 4.3	97.2 ± 6.6	96.9 ± 14.8	91.6 ± 24.4	86.1 ± 9	81 ± 9.2
8	4-HPE	87.3 ± 6.4	74.4 ± 6.4	86.1 ± 5.2	84.3 ± 7.3	40.9 ± 9.7	41.7 ± 9.3	40.8 ± 4.7	44.6 ± 10.4
9	3,4-DHPE	72.2 ± 5.9	59.9 ± 4.5	65.2 ± 5.1	20.2 ± 6.3	13 ± 6.8	21.6 ± 6	20.7 ± 2.8	19.1 ± 3.5
10	4-HBOH	63 ± 9.8	55.3 ± 5.2	62.5 ± 7.4	51.3 ± 14.6	10.2 ± 3.2	14.2 ± 3.5	11.6 ± 6	14.7 ± 4.4
11	3-HPPA	106 ± 8.3	75.6 ± 4	101 ± 3.5	100.6 ± 7.8	125.5 ± 21.7	109.7 ± 29	109.6 ± 10.3	109.8 ± 11.3
12	4-HPPA	96.8 ± 7.3	91.6 ± 6.9	92.4 ± 5.8	100.2 ± 16.8	98.1 ± 7.8	95.8 ± 13.3	88.2 ± 7.7	86.7 ± 17.1
13	3,4-DHPPA	141.6 ± 11.4	99.7 ± 17	78.1 ± 6.3	6.6 ± 20.7	172.8 ± 12.1	168.9 ± 34	126.1 ± 15.5	158.4 ± 59.2
14	DFA	119.8 ± 11.3	92.4 ± 4.5	104.4 ± 4.6	82.4 ± 9.1	111.1 ± 19.6	95.4 ± 28.3	94.8 ± 7.7	95.9 ± 6.1
15	4-HCA	101.1 ± 5.5	93.3 ± 8.2	80.2 ± 6.9	70.9 ± 15.7	94.7 ± 6.8	82.6 ± 8.6	68.6 ± 4.8	56.4 ± 5.4
16	FA	134.5 ± 18.2	119.8 ± 7.4	103.3 ± 8.5	64.1 ± 10.5	115.7 ± 18.7	97 ± 14.7	84.2 ± 6.8	79.6 ± 8.3
17	CA	157.8 ± 11.1	158.4 ± 14.8	126.8 ± 8.1	27.2 ± 6.9	189.3 ± 21.2	180.3 ± 15	150.5 ± 8.9	135.4 ± 14.2
18	3-HPAA	88.5 ± 6.5	80.4 ± 7.8	88 ± 6.9	89.5 ± 28.2	78.5 ± 10	73.7 ± 16	72.9 ± 9	105.9 ± 67
19	4-HPAA	104.9 ± 11.3	94.6 ± 13	76.8 ± 26.7	92.7 ± 251.1	117.2 ± 12.2	138.4 ± 31.1	169.2 ± 73	479.5 ± 974.7
20	3,4-DHPAA	97.6 ± 6	96.7 ± 8.4	68.3 ± 5.1	25.6 ± 5.2	86.1 ± 10	107.3 ± 23.1	91.1 ± 17.8	92.2 ± 116

21	3-HBA	94.8 ± 6.4	88.8 ± 6.2	92.8 ± 4	90.2 ± 11	97.9 ± 7.5	95.4 ± 13	93.1 ± 7.1	94.5 ± 18.2
22	4-HBA	98.6 ± 7.5	96.3 ± 8.7	94.3 ± 16.6	136.4 ± 143.2	96.3 ± 7.1	103.7 ± 15.4	110.2 ± 36.2	268.6 ± 402.3
23	3,4-DHBA	171.2 ± 14.3	152.2 ± 12.6	137.3 ± 6.4	128.6 ± 33.7	213.6 ± 34.4	190.9 ± 23.6	159.5 ± 14.4	171.2 ± 23.2
24	VA	118.3 ± 8.9	99.6 ± 9.2	103.7 ± 7.1	99.5 ± 11.4	113.6 ± 7.5	101.7 ± 20.1	96.2 ± 6.8	99.3 ± 13.1
25	HVA	102 ± 8.4	80.4 ± 7.2	96.5 ± 4.2	81.8 ± 11.9	98.3 ± 21.1	86.7 ± 33.2	77.3 ± 22.8	125.9 ± 216.8
26	HA	38.8 ± 2.9	37.6 ± 3.2	37.4 ± 3.7	35.9 ± 28.5	89.9 ± 6.5	89.1 ± 16.8	87.5 ± 8.4	87.9 ± 16.3

Table E-6. Repeatability error (%) validated for 26 analytes in plasma and brain at four spike levels A-D.

No.	compound	Plasma				Brain			
		A	B	C	D	A	B	C	D
1	RK	3.69	1.69	2.43	3.50	4.67	2.39	5.48	7.01
2	RK-Me	2.70	2.59	2.60	4.46	3.16	1.39	2.67	2.91
3	PhLiAce	1.66	2.81	1.74	2.70	2.68	1.75	1.82	10.18
4	3,4-DHPhLiAce	2.31	0.47	3.95	8.53	2.87	3.01	2.34	4.43
5	VLAce	2.66	1.53	2.11	3.13	2.82	2.91	3.51	3.78
6	VLiAce	1.80	1.82	1.64	3.20	5.13	5.93	4.58	8.18
7	ROH	2.26	2.13	3.13	5.96	4.23	3.97	4.09	7.48
8	4-HPE	2.38	0.84	1.94	2.52	4.70	4.13	3.10	6.74
9	3,4-DHPE	3.89	1.98	2.01	3.43	5.44	4.35	2.85	18.44
10	4-HBOH	2.67	1.48	3.69	5.27	5.77	4.12	8.74	23.30
11	3-HPPA	1.11	5.31	2.84	3.28	3.24	3.59	2.08	5.34
12	4-HPPA	2.28	2.25	2.75	2.95	5.18	4.52	4.02	11.47
13	3,4-DHPPA	2.78	3.35	1.21	18.58	5.47	4.91	9.02	30.80
14	DFA	5.38	1.91	2.82	5.19	3.97	1.81	2.70	3.68
15	4-HCA	2.57	2.40	1.62	3.63	4.83	3.05	2.74	6.36
16	FA	7.92	4.95	4.28	4.72	4.17	2.15	2.22	6.29
17	CA	4.66	3.63	3.33	15.84	6.73	3.85	3.09	6.13
18	3-HPAA	2.61	2.30	2.21	1.99	4.65	4.17	2.06	2.58
19	4-HPAA	2.63	3.23	2.16	2.40	5.42	5.63	4.47	6.21
20	3,4-DHPAA	1.84	5.17	2.20	12.11	3.84	2.94	1.87	4.74
21	3-HBA	2.77	1.84	2.17	4.75	4.19	4.50	3.52	2.58
22	4-HBA	1.82	1.84	1.46	1.23	4.24	2.85	2.04	2.75
23	3,4-DHBA	2.16	1.66	2.13	5.56	4.72	3.98	1.70	5.63
24	VA	2.56	3.31	2.79	3.79	5.66	6.10	1.87	4.72
25	HVA	5.83	2.00	2.47	9.28	7.47	7.01	2.01	1.40
26	HA	1.68	1.95	3.59	7.25	4.29	3.59	2.40	8.22

Table E-7. Matrix effects (%) validated for 26 analytes in plasma and brain at two spike levels B and C.

No.	compound	Plasma		Brain	
		B	C	B	C
1	RK	83.2 ± 18.9	101.9 ± 2.1	88.6 ± 20.2	107.4 ± 2.7
2	RK-Me	80.9 ± 3	82.3 ± 3.6	95.1 ± 4.3	88.2 ± 1.8
3	PhLiAce	94.2 ± 3.1	98.8 ± 3.5	108.1 ± 5.1	105.8 ± 3.2
4	3,4-DHPhLiAce	165 ± 15.8	227.7 ± 28.1	165.3 ± 14.3	218.3 ± 27.1
5	VLAce	107.3 ± 3.9	113.9 ± 2.6	113.1 ± 3.2	114 ± 2.3
6	VLiAce	117.7 ± 6.4	128.9 ± 5.6	132.9 ± 6.2	139.9 ± 4.7
7	ROH	100.2 ± 4.5	102.9 ± 4	113.7 ± 4.6	110.6 ± 3.3
8	4-HPE	89.5 ± 4.5	87.2 ± 7.5	83.5 ± 2.1	74.2 ± 2.3
9	3,4-DHPE	120.6 ± 5.9	136.1 ± 32	93.6 ± 4.4	102.5 ± 10.9
10	4-HBOH	93.9 ± 3.8	94.7 ± 9.4	95.8 ± 2	94.7 ± 3.2
11	3-HPPA	103.8 ± 4.1	107.2 ± 4.7	115.1 ± 3.7	120 ± 3.4
12	4-HPPA	105.6 ± 4.3	111.9 ± 7.3	115.8 ± 4.6	120.2 ± 6.4
13	3,4-DHPPA	227.5 ± 41.9	275.5 ± 35.7	235.6 ± 42.7	280 ± 41
14	DFA	123.7 ± 7.7	141.7 ± 11.2	114.8 ± 5.5	123.1 ± 11.5
15	4-HCA	123.9 ± 7	146.4 ± 11.3	145.4 ± 7.8	162.7 ± 10.6
16	FA	194.1 ± 25	254.9 ± 44.5	172.9 ± 22	212.3 ± 37.2
17	CA	401.6 ± 108.6	710.4 ± 125.3	493.3 ± 133.2	738.5 ± 129
18	3-HPAA	89.4 ± 3.1	91.3 ± 6.1	82.7 ± 2	78.9 ± 4.1
19	4-HPAA	66.3 ± 5.6	60 ± 10.8	61.4 ± 6.4	44.8 ± 15.2
20	3,4-DHPAA	260.2 ± 64.7	414.5 ± 144.7	197.9 ± 51.5	278.1 ± 98.7
21	3-HBA	99.5 ± 4.1	102.8 ± 5.3	112.7 ± 3.4	111.1 ± 2
22	4-HBA	93.5 ± 3.8	93.5 ± 15.3	101.4 ± 4.4	93.2 ± 14.8
23	3,4-DHBA	200.5 ± 32.3	237.3 ± 56.2	264.6 ± 42.6	322 ± 70.7
24	VA	119 ± 6.7	131.8 ± 11.5	124 ± 6.6	137.4 ± 9
25	HVA	121.6 ± 11.4	143.6 ± 16.7	85.1 ± 8.3	89.6 ± 14.4
26	HA	99 ± 3.3	98.1 ± 9.4	116.1 ± 3.7	113.5 ± 2.2

Table E-8. Recovery (%) validated for 26 analytes in plasma and brain at two spike levels B and C.

No.	compound	Plasma		Brain	
		B	C	B	C
1	RK	94.7 ± 3.1	88.8 ± 4.6	71.5 ± 4.7	69.9 ± 5.6
2	RK-Me	70.1 ± 6.5	63.5 ± 8.3	61.7 ± 7.5	65.4 ± 4.7
3	PhLiAce	34.3 ± 10.7	33.9 ± 8.5	47.5 ± 8.1	55.2 ± 5.1
4	3,4-DHPhLiAce	82.8 ± 4.7	69.8 ± 6.5	73.5 ± 5	67.8 ± 6.3
5	VLAce	91.7 ± 3.7	86.6 ± 6.3	70.5 ± 5.2	70.8 ± 4.2
6	VLiAce	90.2 ± 4.7	83.2 ± 7.2	69 ± 5	66.4 ± 4.8
7	ROH	93 ± 4.2	90.8 ± 5.8	67.3 ± 3.7	61.8 ± 4.3
8	4-HPE	93.5 ± 4.8	88.6 ± 8.7	46 ± 2.8	43.8 ± 3.2
9	3,4-DHPE	64.9 ± 2.4	53.5 ± 12.4	24.2 ± 2.5	20.1 ± 2.4
10	4-HBOH	62.3 ± 6.7	62.5 ± 8.7	12.7 ± 1.2	11.6 ± 3.3
11	3-HPPA	92.8 ± 3.6	89.7 ± 6.1	75.6 ± 4.3	72.5 ± 4.5
12	4-HPPA	93.3 ± 3.2	88.5 ± 7.5	72.3 ± 3.9	69.2 ± 6.4
13	3,4-DHPPA	68.1 ± 4.9	50.5 ± 4.5	75.9 ± 2.9	69.9 ± 9.5
14	DFA	92.2 ± 4.5	85.2 ± 6	76.7 ± 3.4	74.1 ± 6.3
15	4-HCA	87.4 ± 3	92.2 ± 6.7	55.6 ± 5.4	62.3 ± 4.5
16	FA	78.9 ± 5	78.8 ± 8.5	60.6 ± 6.3	64.3 ± 6
17	CA	74.6 ± 4.5	53.4 ± 3.7	57.6 ± 5.5	54.1 ± 3.8
18	3-HPAA	93.6 ± 3.3	93.3 ± 6.7	76.1 ± 2.8	81.5 ± 3.9
19	4-HPAA	102.2 ± 3.1	101.3 ± 7.3	88.1 ± 6.1	95.8 ± 5
20	3,4-DHPAA	56.1 ± 2.4	39.5 ± 4.5	69 ± 5.8	78.1 ± 7.1
21	3-HBA	95.8 ± 4.2	90.5 ± 5.9	73.4 ± 4.2	74.8 ± 5.8
22	4-HBA	99.9 ± 3.3	101.9 ± 7.8	80.3 ± 5.6	93 ± 3.9
23	3,4-DHBA	92.9 ± 3.8	89.7 ± 10.7	69.6 ± 4.8	68.8 ± 6.5
24	VA	97.2 ± 4.2	91.8 ± 7.8	77 ± 4.1	72.6 ± 3.7
25	HVA	85.9 ± 3.1	85 ± 5.5	82.8 ± 2.8	91 ± 5.1
26	HA	40.5 ± 2	41.7 ± 4	63.6 ± 2.5	62.7 ± 4.3

Table E-9. Processing efficiency (%) validated for 26 analytes in plasma and brain at two spike levels B and C.

No.	compound	Plasma		Brain	
		B	C	B	C
1	RK	78.8 ± 17.9	90.3 ± 4.5	63.2 ± 14.7	73.9 ± 6.1
2	RK-Me	56.7 ± 4.9	52.1 ± 6.5	58.7 ± 6.8	57.7 ± 4
3	PhLiAce	32.3 ± 10.1	33.2 ± 8.4	51.3 ± 8.4	58.4 ± 5.1
4	3,4-DHPhLiAce	136.6 ± 12.5	158.8 ± 24.3	121.6 ± 13	147.9 ± 22.5
5	VLAce	98.4 ± 2.6	98.5 ± 7.1	79.7 ± 5.7	80.6 ± 4.8
6	VLiAce	106.2 ± 5.2	107.1 ± 9.5	91.7 ± 7	92.9 ± 7.3
7	ROH	93.1 ± 3.3	93.4 ± 6.2	76.3 ± 4	67.3 ± 5.3
8	4-HPE	83.7 ± 2.9	77.1 ± 4.9	38.3 ± 2.4	32.1 ± 2.5
9	3,4-DHPE	78.2 ± 4.5	72.4 ± 6.7	22.6 ± 2.6	20.4 ± 2.3
10	4-HBOH	58.3 ± 6	58.5 ± 6.2	12 ± 1.1	10.3 ± 3.2
11	3-HPPA	96.3 ± 2.7	96 ± 6.2	87 ± 4.8	86.7 ± 5.7
12	4-HPPA	98.4 ± 2.9	98.3 ± 7.4	83.7 ± 4.3	82.9 ± 7.3
13	3,4-DHPPA	154.8 ± 29.7	137.4 ± 20.9	178.8 ± 33	195.9 ± 32.7
14	DFA	114 ± 6	120.3 ± 12.5	88 ± 5.5	91.2 ± 9.2
15	4-HCA	108 ± 5.8	133.6 ± 12.8	80.7 ± 8.6	100.9 ± 9.9
16	FA	153 ± 20.9	199.7 ± 40.1	104.7 ± 16.8	136.4 ± 26
17	CA	299.4 ± 82.2	379.2 ± 70.1	283.6 ± 80.8	396.7 ± 74.4
18	3-HPAA	83.3 ± 1.6	83 ± 8	60.6 ± 2.2	55 ± 5.7
19	4-HPAA	68.9 ± 6.8	64 ± 19.8	43.7 ± 9.6	26.6 ± 21
20	3,4-DHPAA	145.9 ± 36.2	162.2 ± 56.1	127.9 ± 32.2	166.4 ± 62.5
21	3-HBA	95.3 ± 2.7	92.6 ± 4.4	82.4 ± 4.4	81.3 ± 6.9
22	4-HBA	93.4 ± 2.8	99.3 ± 17.9	69.1 ± 9.1	62.6 ± 21.5
23	3,4-DHBA	186.2 ± 29.2	212.3 ± 44.9	183.2 ± 30.6	215.8 ± 47.2
24	VA	115.6 ± 6	120.7 ± 11.2	95.3 ± 6.1	98.4 ± 8.3
25	HVA	104.3 ± 9.5	121.5 ± 15.7	62.7 ± 6.2	58.4 ± 17.9
26	HA	38.7 ± 1.7	34.2 ± 2.8	73.6 ± 2.2	69.8 ± 5

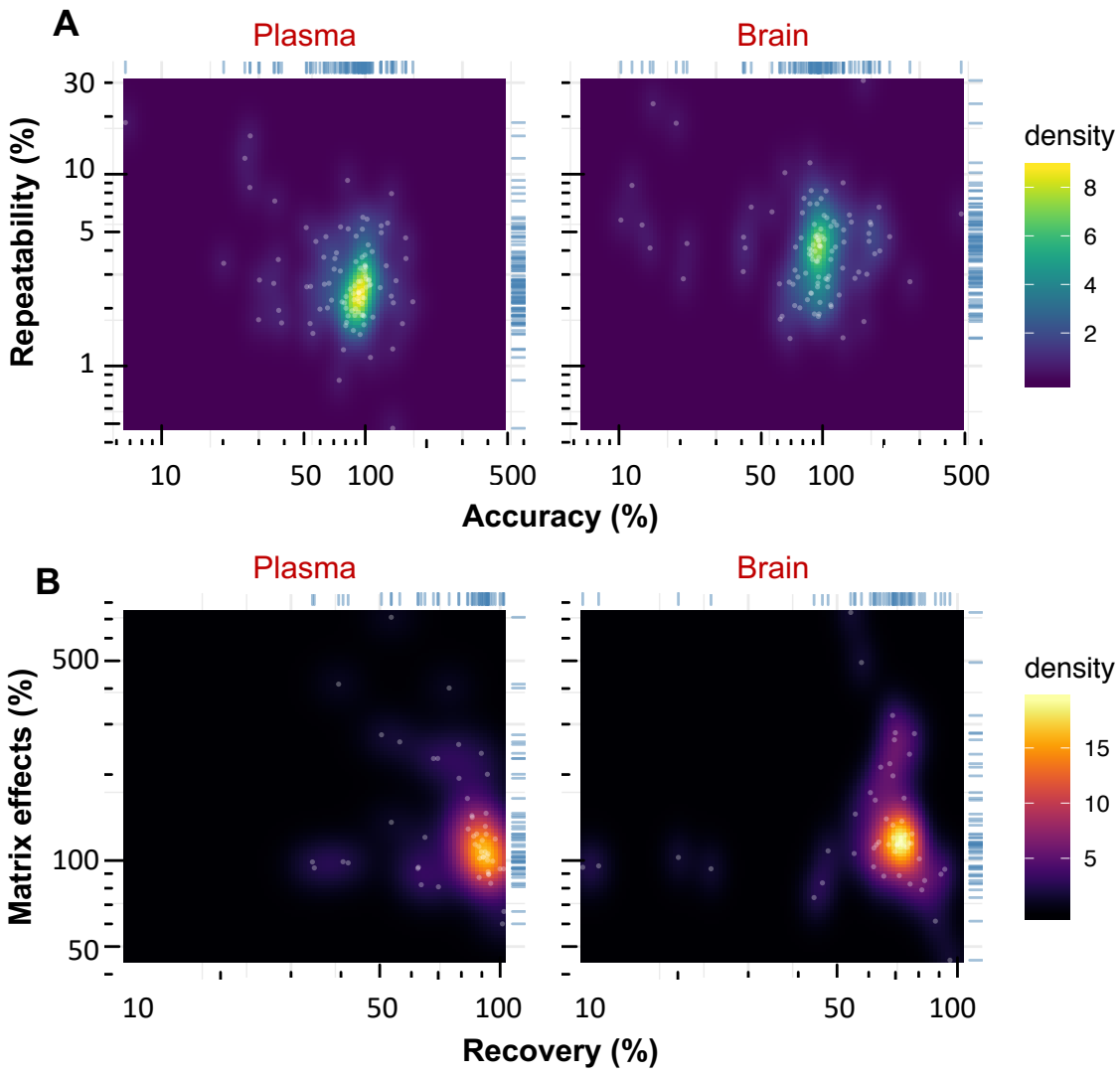


Figure E-4. Two-dimensional density plots of (**A**) accuracy vs. repeatability, and (**B**) recovery vs. matrix effects. All compounds and spike levels validated are displayed. The marginal barcode-like plots present the corresponding one-dimensional data distribution. Note that axes are logarithmically scaled.

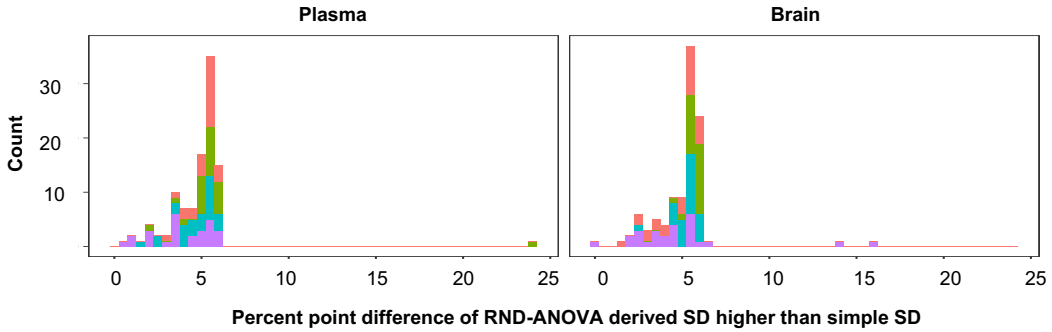


Figure E-5. Distribution of numeric difference of standard deviation (SD) of measured spiked concentration in biomatrices calculated by random effects analysis of variance (RND-ANOVA) vs. (higher than) otherwise not used (using the ordinary SD formula instead). The SD of measured spiked concentration in biomatrices are the major constituent of accuracy variability in most cases (refer to prior chapter **Figure D-4**).

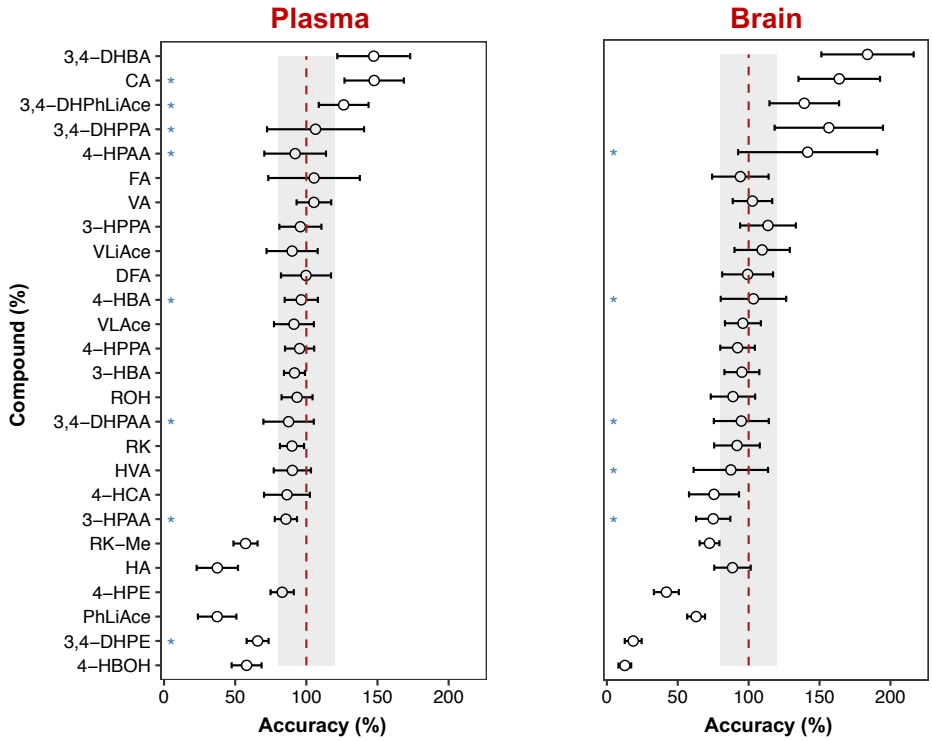


Figure E-6. Accuracy inference across linear dynamic range (LDR) using random effects analysis of variance (RND-ANOVA). The inference was made based on three or four spike levels (level A, 2000 ng/mL; B, 1000 ng/mL; C, 150 ng/mL; and D, 15 ng/mL), with five replicates per spike level. For compounds noted with blue stars, the level D spike concentration was not included in calculation, considering the large data volatility or aberrance due to blank or other interference at this level. The shaded area denotes 80~120 % range.

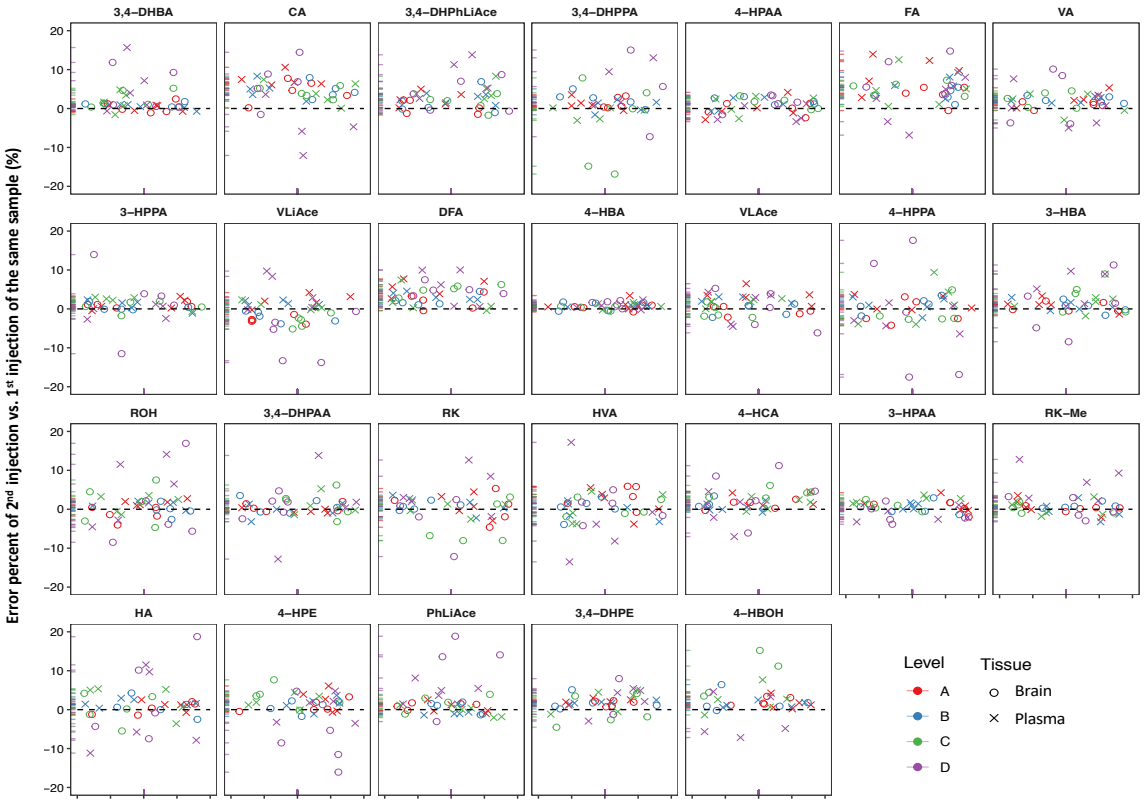


Figure E-7. The peak area error percent of the 2nd relative to the 1st injection (spaced by *ca* 10 hours) of the same quality control samples (QCS) of plasma and brain in the validation experiment. The error percent was calculated as (Peak Area $_{\text{2nd injection}}$ − Area $_{\text{1st injection.}}$) / Area $_{\text{1st injection}}$ × 100%. The barcode-like plot on the left inside each faceted plot show the data distribution of corresponding spike concentration. X-axis does not hold practical meaning; it's intended only for display convenience with point scatterings. A minor number of outliers beyond the applied scales are not displayed. Dots' position reflected compound degradation and instrumental precision. All compounds presented unperceivable degradation over 10 hours in biomatrices; though for some compounds such as DFA, FA and CA, the 2nd injections showed almost consistently higher response than the 1st injections, possibly due to instrumental drifting (mostly less than 10%). Dot dispersion is calculated as repeatability on a level-specific manner. For spike levels, A (2000 ng/mL), B (1000 ng/mL), C (150 ng/mL) and D (15 ng/mL), with concentration shown in the final processed samples before injections. Autosampler was maintained at 4°C.

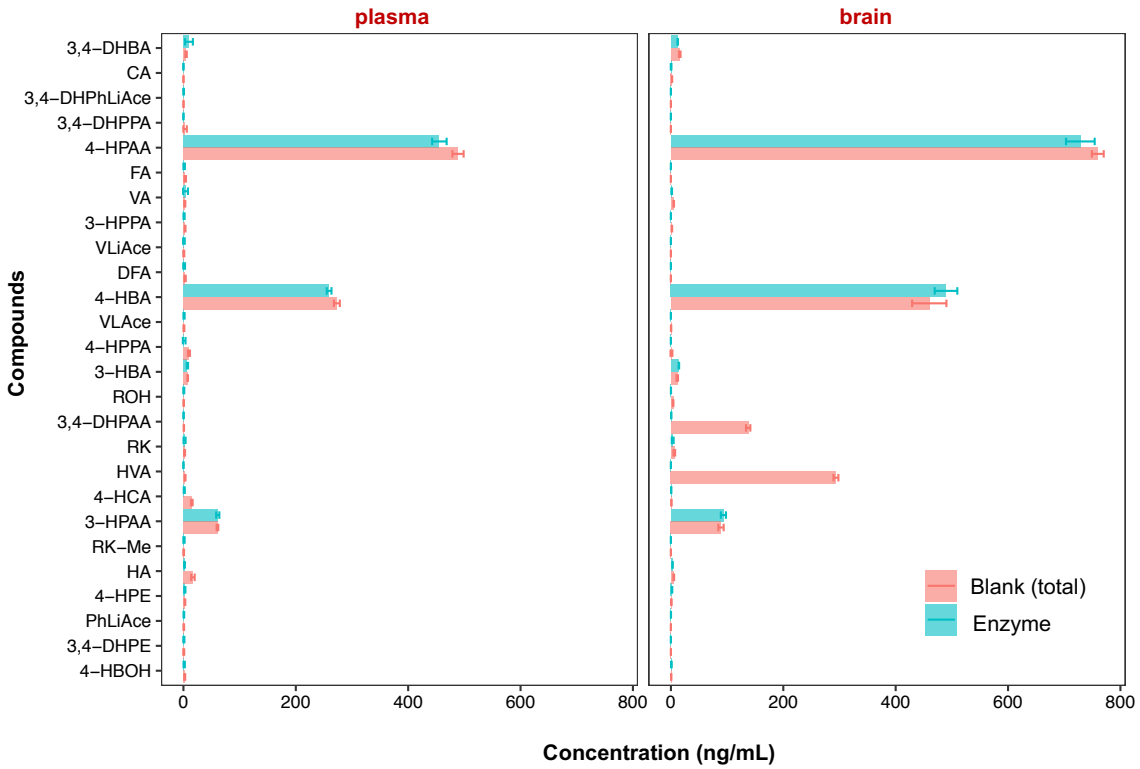


Figure E-8. Analytes' concentration in the *total blank* and the *exogenous* proportion originating from β-glucuronidase enzyme solution (from limpets or *Patella vulgata*). The concentration difference between total blank and enzyme-derived amount is the *endogenous* quantity from mice tissues. The concentration presented here is the level in final processed sample before injection (so as for convenience of comparison with spike concentrations).

Note:

1) *total blank* = *exogenous* quantity from enzyme + *endogenous* quantity from mice tissues (the vehicle control)

2) The amount of enzyme solution (2000 U) used for processing brain samples was twice that of plasma.

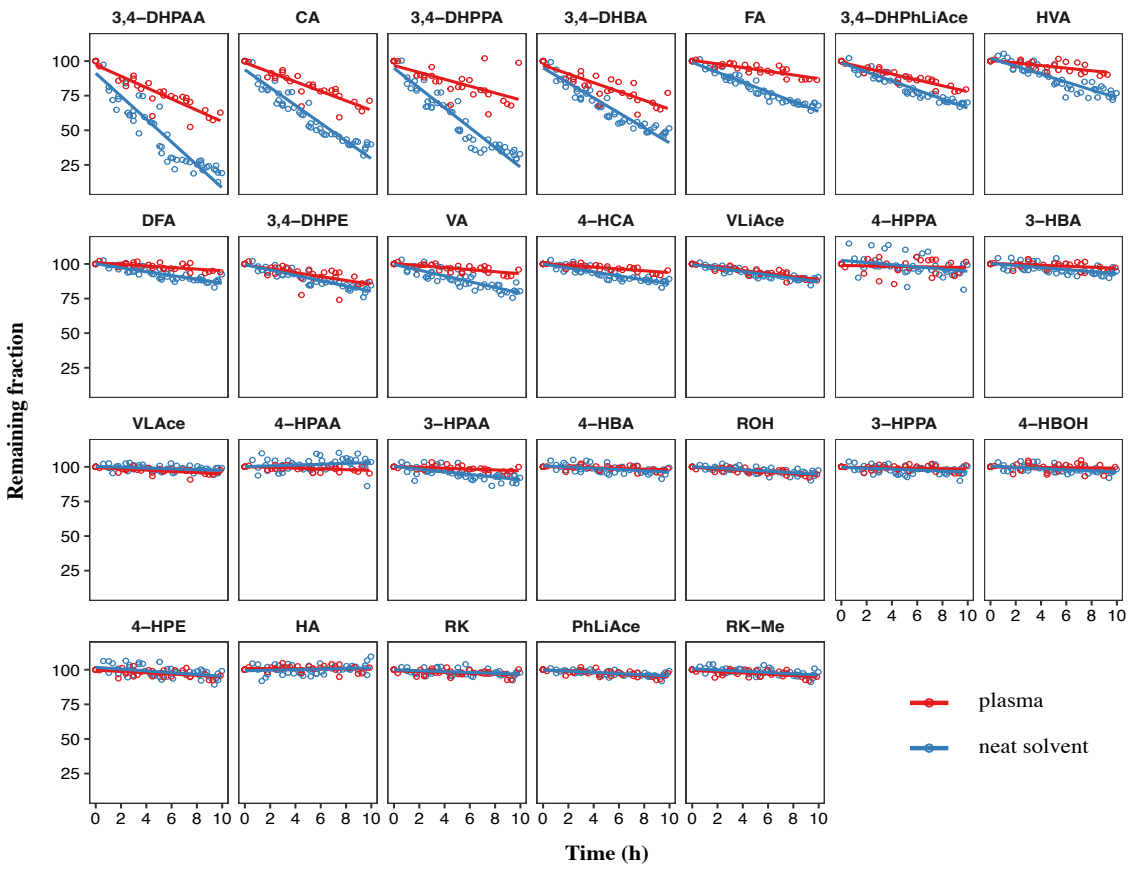


Figure E-9. Compound degradation profile across ten hours in pure solvent and plasma matrices in a separate study. This study was independent of all prior experiments aforementioned. Compounds are displayed in decreasing order of degradation rate. Compound in pure solvent (60 % methanol with 0.1% formic acid) showed similar degradation profile as in prior study **Figure E-3**. Compounds liable in pure solvent exhibited noted degradation in plasma matrices, which somewhat disagreed from the unperceived degradation shown in the validation study (as shown in **Figure E-7**), possibly due to dissimilarities in the involved biomatrices from different lots or longer observation time. However, both studies congruently demonstrated improved compound stability in the biomatrices than in pure solvent.

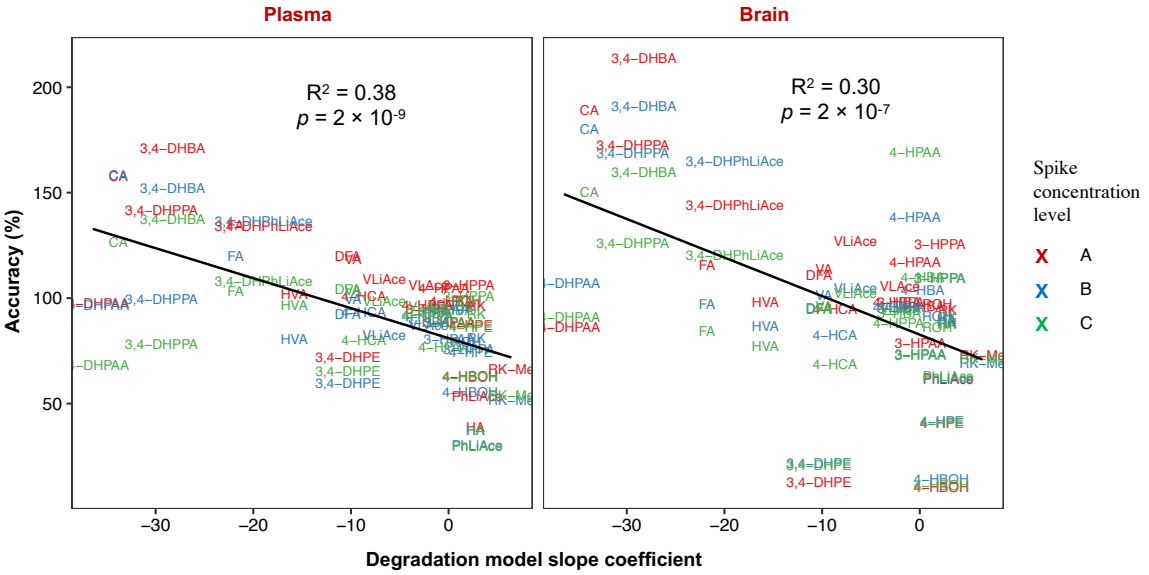


Figure E-10. Correlation of accuracy with compound degradation in pure solvent. The slope coefficient of zero-order kinetic model noted on the x-axis corresponds to the percent loss per hour. Depending on liability, compound degradation in the calibration work solution prepared in *pure solvent* (60% methanol with 0.1% formic acid) could cause systematic error to various extent for quantification in *biomatrices*, leading to numerically higher accuracy (e.g. CA and 3, 4-DHBA above 150%).

APPENDIX F. URL's (R SCRIPT & SHINY APP)

❖ **Nightshade leafy phytochemical quantification**

https://yuanbofaith.github.io/NSleaf_PhytochemQqQ/

❖ **Free amino acid analysis in African indigenous vegetables and classification prediction using machine learning**

https://yuanbofaith.github.io/AfricanVegetables_AminoAcids/

R Shiny app: *https://boyuan.shinyapps.io/AIV_Classifier/*

❖**Nightshade glycoalkaloid rapid screening**

https://yuanbofaith.github.io/Solanum_alkaloid_in-source-fragmentation_MSMS/

❖**Raspberry ketone metabolomics study (1)** *(plasma and brain analysis)*

https://yuanbofaith.github.io/RK_LCMS/

❖**Raspberry ketone metabolomics study (2)** *(adipose tissue cleanup using EMR-lipid cleanup sorbent in 96-well plate)*

https://yuanbofaith.github.io/RK_adipose_QuEChERS_EMR/index.html

❖ **Logistic regression vs. Gaussian discriminant analysis comparison**

R Shiny app:
https://boyuan.shinyapps.io/Logistic_and_Gaussian_Discriminant_Analysis/

❖ **Lemon juice authentication vs. adulteration classification prediction** *(a collaboration project, all data collected and analyzed by Weiting Lyu)*

https://yuanbofaith.github.io/Lemon_Juice_Classification2/index.html

R Shiny app: *https://boyuan.shinyapps.io/LemonClassification/*

❖**Catnip quality control visualization tool**

(a collaboration project with Erik Gomes et al.)

https://boyuan.shinyapps.io/CatnipQC/

https://yuanbofaith.github.io/Catnip_ShinyVisualization_RawScript/

❖**R visualization mini-gallery**

https://yuanbofaith.github.io/Rvision/

❖***Protag*: searching tagged peptides based on MALDI-TOF MS peptide fingerprint**

(a self-motivated project inspired by an internship project)

- *R Shiny app*: *https://boyuan.shinyapps.io/protag/*
- R package downloadable from CRAN *https://cran.r-project.org/web/packages/protag/index.html*
- Package user guide: *https://rpubs.com/Boyuan/guide_protag*