Comparison of Flow Injection MS, NMR, and DNA Sequencing: Methods for Identification and Authentication of Black Cohosh (*Actaea racemosa*)

Authors James Harnly[1], Pei Chen[1], Jianghao Sun[1], Huilian Huang[1], Kimberly L. Colson[2], Jimmy Yuk[2], Joe-Ann H. McCoy[3], Danica T. Harbaugh Reynaud[4], Peter B. Harrington[5], Edward J. Fletcher[6]

Affiliations The affiliations are listed at the end of the article

Key words
- *Actaea racemose*
- Ranunculaceae
- NMR
- FIMS
- DNA sequencing

Abstract

Flow injection mass spectrometry and proton nuclear magnetic resonance spectrometry, two metabolic fingerprinting methods, and DNA sequencing were used to identify and authenticate *Actaea* species. Initially, samples of *Actaea racemosa* from a single source were distinguished from other *Actaea* species based on principal component analysis and soft independent modeling of class analogies of flow injection mass spectrometry and proton nuclear magnetic resonance spectrometry metabolic fingerprints. The chemometric results for flow injection mass spectrometry and proton nuclear magnetic resonance spectrometry agreed well and showed similar agreement throughout the study. DNA sequencing using DNA sequence data from two independent gene regions confirmed the metabolic fingerprinting results. Differences were observed between *A. racemosa* samples from four different sources, although the variance within species was still significantly less than the variance between species. A model based on the combined *A. racemosa* samples from the four sources consistently permitted distinction between species. Additionally, the combined *A. racemosa* samples were distinguishable from commercial root samples and from commercial supplements in tablet, capsule, or liquid form. DNA sequencing verified the lack of authenticity of the commercial roots but was unsuccessful in characterizing many of the supplements due to the lack of available DNA.

received January 29, 2015
revised Sept. 2, 2015
accepted Sept. 3, 2015

Bibliography
DOI http://dx.doi.org/10.1055/s-0035-1558113
Published online December 21, 2015
Planta Med 2016; 82: 250–262
© Georg Thieme Verlag KG Stuttgart · New York ·
ISSN 0032-0943

Correspondence
James M. Harnly
Food Composition and Methods Development Laboratory
Beltsville Human Nutrition Research Center
Agricultural Research Service
U.S. Department of Agriculture
10300 Baltimore Ave,
Building 161
Beltsville, MD 20705
USA
Phone: + 1 301 504 85 69
Fax: + 1 301 504 83 14
james.harnly@ars.usda.gov

Introduction

This study describes the application of two chemical fingerprinting methods and a genetic sequencing method for the authentication of black cohosh [*Actaea racemosa* L. (Ranunculaceae)], one of the top ten selling dietary supplements in the U.S. [1]. The three methods include flow injection mass spectrometry (FIMS), proton nuclear magnetic resonance (^{1}H-NMR), and DNA sequencing using universal primer regions. *A. racemosa* is a particularly appropriate target botanical for this study as its increasing commercial demand has resulted in frequent adulteration and substitution for economic purposes.

Identification and authentication of botanical materials is a challenging task due to their complex chemistry, phylogeny, and numerous material forms which include dried plant material, ground powder, liquid extracts, and dried extracts. The gold standard for taxonomy is based on morphological characteristics from a whole plant, primarily of the flower and/or fruit by an expert botanist [2]. Even this approach, however, has its limitations, as plant material received for commercial production may not match the original voucher specimen or may contain organs or plant parts that aren't typically used for taxonomic identification (i.e., roots). Macroscopic and microscopic methods have also been used extensively, but as raw plant materials are ground and extracted, qualitative methods become less applicable. Ground botanical materials lose their morphological characteristics but can retain their genetic identity. As a result, DNA sequencing is becoming an increasingly routine and affordable method for taxonomic discrimination of ground plant materials.

Identification of extracted botanical materials can be highly problematic. Extraction targets broad groups of compounds based on the polarity of the solvent and is frequently used for enrichment of desirable components and/or removal of undesirable components. Consequently, extraction will alter the chemical composition of a botanical as compared to the original solid material. Extrac-

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

tion can also be problematic for DNA sequencing. The availability of sufficient lengths of DNA sequences is highly dependent on the extraction technique and the DNA sequencing method.

One of the most comprehensive approaches to botanical authentication based on chemical composition is metabolic fingerprinting. Hall [3] defined metabolic fingerprinting as "High throughput qualitative screening of the metabolic composition of an organism or tissue with the primary aim of sample comparison and discrimination analysis. Generally no attempt is initially made to identify the metabolites present. All steps from sample preparation, separation, and detection should be rapid and as simple as is feasible". Chromatographic and spectral (with no prior separation) fingerprinting can meet Hall's definition. Both approaches generate complex fingerprints that may require multivariate analysis for discrimination. A general approach to the validation of botanical identification methods using multivariate analysis has been described by AOAC International [4,5].

Numerous methods have been developed and utilized for the authentication of *A. racemosa* [6–9], although none of them include chemical fingerprinting methods. The most commonly used methods rely on the identification and quantification of specific metabolites (chemical markers), an approach defined as metabolic profiling. Ma et al. [6] described a method based on high-performance liquid chromatography coupled with electrospray ionization/mass spectrometry (HPLC-ESI/MS) to identify 15 chemical markers. Qiu et al. [7] described a method using NMR for the measurement of triterpenes in *A. racemosa*. The two-dimensional plots obtained from heteronuclear single-quantum coherence (HSQC) measurements (combining ^{1}H-NMR and ^{13}C-NMR) were used to generate patterns for distinguishing the various *Actaea* species. Neither of these methods are comprehensive because they depended solely on identified components and ignore the bulk of the components observed in the MS and NMR spectra. Furthermore, neither study validated the markers by measuring identification rates.

In previous studies, Harnly's lab [10–15] has demonstrated the use of spectral fingerprints obtained by IR, NIR, FIMS (with both positive and negative ionization), and UV spectrometry for characterizing botanicals (bitter orange, black cohosh, *Ginkgo*, and ginseng) and food plants and materials (broccoli, dry beans, grapefruit, skim milk powder). Spectral fingerprinting allowed for the discrimination between species, growing locations, growing conditions, and processing. FIMS proved to be a particularly useful method [14–16]. Whereas chromatographic methods require special care for retention time alignment, normal mass calibration procedures for nominal and high-resolution FIMS provide suitably stable fingerprints. The advantages of FIMS are the excellent sensitivities and potential for identifying discriminating components (especially with high-resolution FIMS) from the variable loadings obtained from the chemometric models. However, ion counts are highly dependent on the ionization process and reproducibility between experiments on the same instrument and between different instrument designs can be poor. This generally necessitates a batch type of operation, i.e., fingerprints can only be compared between new and reference samples analyzed under the same conditions and at the same time. This process rapidly depletes the supply of reference samples.

NMR, by comparison, has not been used as frequently for spectral fingerprinting, although several excellent studies have been reported [17–21]. In general, NMR may be considered less sensitive than MS, more expensive with respect to cost and consumables, and requires an extra sample preparation step of drying and reconstitution in a deuterated solvent. However, its inherent stability makes it an excellent candidate for fingerprinting, allowing the possibility of simply comparing spectra of new materials with archived spectra to determine authenticity. Like MS, the NMR spectra provide the added possibility of identifying specific components that provide sample discrimination. In addition, the uniform response of ^{1}H-NMR facilitates calibration and allows for the quantification of compounds.

Fingerprint comparisons of FIMS and NMR spectra are assisted by multivariate analysis methods. There are two general approaches: modeling (a soft method) and classification (a hard method). Soft models, such as the soft independent modeling of class analogy (SIMCA) [22] and the fuzzy optimal associative memory (FOAM) [23], fit a model to a single class. They are also referred to as one-class classifiers because they model the similarities among features of the spectra within just a single class. As a result, they are a useful tool for authentication. A set of authentic objects is used to construct a model that is used to judge whether the test samples are authentic or not. Compared to classification methods, modeling methods tend to have less discriminating power. However, they have the advantage of being able to reject novel spectral fingerprints with characteristics not included in the model.

Hard models, or classification methods, such as the partial least squares-discriminant analysis (PLS-DA) [24] and the fuzzy rule-building expert system (FuRES) [25], require identification of every class to be considered in the model and force an unknown sample into one of the classes. PLS-DA builds harder models as the number of components increase. FuRES is a soft classification method by its fuzzy constraints. Classification works well when the classes of samples are well known and potential adulterants are known ahead of time. Classification models will typically misclassify new fingerprints that do not belong to any of the classes, i.e., have features not incorporated into the model.

A third method that can be used for authentication is DNA sequencing, which uses universal gene regions to identify organisms [26]. Like metabolic fingerprinting, the pattern of the sequence is matched with reference sequences to authenticate the botanical identity. DNA sequencing provides a genotypic fingerprint, not a metabolic fingerprint. The "DNA barcode", which the DNA sequence is often referred as, is complementary to metabolic fingerprinting, allowing for the identification of species and, in some cases, subspecies or variety. It is not appropriate for discriminating between plant parts or the influence of the environment on metabolite expression. As the cost of DNA sequencing drops, it is becoming the method of choice for the authentication of raw animal and plant materials. For finished product supplements, the method may be inappropriate as the absence of DNA, or presence of low quality DNA, is problematic [27]. However, new sequencing methods are being developed with more species-specific primers that target short fragments of DNA as would be expected in finished products [28].

The purpose of this study was to examine the ability of NMR and FIMS, in combination with multivariate analysis, to discriminate between *Actaea* species, between raw materials and commercial products, and to compare spectral classification with DNA sequencing. First, FIMS and ^{1}H-NMR fingerprints of *Actaea* species obtained from a single source were compared using principal component analysis (PCA) and SIMCA. Identification of the species was confirmed using DNA sequencing based on two independent nuclear ribosomal and chloroplast gene regions as validated for *Actaea* [29]. Samples of authentic *Actaea* species from

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

Table 1 American Herbal Pharmacopoeia samples.

ID #	Species	Part	Form	Source	Location	nrDNA	Mix	Hybrid	cpDNA	Mix
American Herbal Pharmacopoeia										
BCR01	*A. racemosa*	r/r	powder	AHP	North Carolina	*A. racemosa*	yes	n/a	*A. racemosa*	no
BCR02	*A. pachypoda*	r/r	powder	AHP	North Carolina	*A. pachypoda*	no	no	*A. pachypoda*	no
BCR03	*A. racemosa*	r/r	powder	AHP	China	*A. pachypoda*	no	no	pqs	n/a
BCR04	*A. racemosa*	r/r	powder	AHP	New Jersey	*A. racemosa*	little	no	*A. racemosa*	no
BCR05	*A. pachypoda*	r/r	powder	AHP	North Carolina	*A. pachypoda*	no	no	*A. pachypoda*	no
BCR06	*A. racemosa*	r/r	powder	AHP	NA	*A. racemosa*	no	maybe	*A. racemosa*	no
BCR07	*A. racemosa*	r/r	powder	AHP	North Carolina	*A. racemosa*	little	maybe	No ID	yes
BCR08	*A. racemosa*	r/r	powder	AHP	North Carolina	*A. racemosa*	little	maybe	No ID	yes
BCR09	*A. racemosa*	r/r	powder	AHP	North Carolina	*A. racemosa*	no	maybe	*A. racemosa*	no
BCR10	*A. podocarpa*	r/r	powder	AHP	North Carolina	*A. podocarpa*	no	no	*A. podocarpa*	no
BCR11	*A. podocarpa*	r/r	powder	AHP	North Carolina	*A. podocarpa*	no	no	*A. podocarpa*	no
BCR12	*A. podocarpa*	r/r	powder	AHP	North Carolina	*A. podocarpa*	no	no	*A. podocarpa*	no
BCR13	*A. cimcifuga*	r/r	powder	AHP	China	*A. dahurica*	no	no	*A. dahurica*	no
BCR14	*A. rubra*	r/r	powder	AHP	Quebec, Canada	*A. rubra*	no	yes	*A. pachypoda*	no
BCR15	*A. rubra*	r/r	powder	AHP	Oregon	*A. rubra*	no	no	*A. rubra*	no
BCR16	*A. racemosa*	r/r	powder	AHP	commercial	*A. racemosa*	no	maybe	*A. racemosa*	no
BCR17	*A. racemosa*	r/r	powder	AHP	commercial	*A. racemosa*	no	maybe	*A. racemosa*	no
BCR18	*A. cimcifuga*	r/r	powder	AHP	commercial	*Erotium*	yes	n/a	No ID	n/a
BCR19	*A. cimcifuga*	r/r	powder	AHP	China	*Erotium*	yes	n/a	No ID	yes
BCR20	*A. cimcifuga*	r/r	powder	AHP	China	*Erotium*	yes	n/a	No ID	yes
BCR21	*A. cimcifuga*	r/r	powder	AHP	commercial	*Erotium*	yes	n/a	No ID	n/a
BCR22	*A. cimcifuga*	r/r	powder	AHP	commercial	*A. dahurica*	yes	n/a	*A. dahurica*	no
BCR23	*A. cimcifuga*	r/r	powder	AHP	China	*A. dahurica*	yes	n/a	*A. dahurica*	no
BCR24	*A. cimcifuga*	r/r	powder	AHP	China	*A. dahurica*	yes	n/a	*A. dahurica*	no

nrDNA: nuclear ribosomal DNA; cpDNA: chloroplast DNA; r/r: roots/rhizomes; n/a: not analyzed; pqs: poor quality sequence; NA: not available

Table 2 National Institute of Standards and Technology and Strategic Sources samples.

ID #	Species	Part	Form	Source	Location	nrDNA	Mix	Hybrid	cpDNA	Mix
National Institute of Standards and Technology										
SRM3295	*A. racemosa*	r/r	powder	NIST	NA	*A. racemosa*	yes	no	*A. racemosa*	yes
SRM3296	*A. racemosa*	r/r	powder	NIST	NA					
SRM3297	*A. racemosa*	r/r	powder	NIST	NA					
SRM3298	*A. racemosa*	r/r	powder	NIST	NA					
Strategic Sources										
SS01	*A. racemosa*	r/r	r/r	SS	Madison, AL	*A. racemosa*	no	maybe	*A. racemosa*	no
SS02	*A. racemosa*	r/r	r/r	SS	Bell, KT	*A. racemosa*	no	maybe	*A. racemosa*	no
SS03	*A. racemosa*	r/r	r/r	SS	Logan, WV	*A. racemosa*	yes	n/a	*A. racemosa*	no
SS04	*A. racemosa*	r/r	r/r	SS	Carter, MO	*A. racemosa*	yes	n/a	*A. racemosa*	no
SS05	*A. racemosa*	r/r	r/r	SS	Washington, MO	*A. racemosa*	yes	maybe	*A. racemosa*	no
SS06	*A. racemosa*	r/r	r/r	SS	Clay, KT	*A. racemosa*	no	no	*A. racemosa*	no
SS07	*A. racemosa*	r/r	r/r	SS	Pike, KT	*A. racemosa*	yes	maybe	*A. racemosa*	no

nrDNA: nuclear ribosomal DNA; cpDNA: chloroplast DNA; r/r: roots/rhizomes; n/a: not analyzed; NA: not available

multiple sources were then examined using FIMS, ^{1}H-NMR, and DNA sequencing and compared to the supplier's identification. Similarly, metabolic fingerprints and DNA sequences were used to compare authentic *A. racemosa* samples with both commercially available root samples and with commercially available supplements. Four multivariate methods (SIMCA FOAM, PLS-DA, and FuRES) were used to evaluate the data in this study.

Results and Discussion

MS and NMR spectra were acquired for the *Actaea* species listed in **Tables 1–4**. Typical spectra for *A. racemosa* are given in **Fig. 1**. The FIMS spectra were acquired with flow injection (no separation) for *m/z* 150 to 1500. The NMR spectra were acquired from −3.0 to 16.0 ppm, but only the region from 0.5 to 9.0 ppm was used in this study. The top spectra of **Fig. 1** have not undergone any preprocessing. The bottom plots show both spectra after normalization to unit vector length and autoscaling with respect to the *A. racemosa* spectra. The NMR spectra had the DMSO solvent peak at 2.53 ppm removed before the preprocessing steps. The solvent peak can potentially introduce added undesirable variance to the data and may confound classification and modeling, although this was not the case (data not shown).

The initial goal was to compare the ability of FIMS and NMR to discriminate between species. Because the PCA of the spectra for all the samples in **Tables 1–4** produced very complex score plots that were difficult to interpret, a reductionist approach was deemed necessary for this phase of the investigation. Consequently, only spectra for *A. racemosa* and four other *Actaea* spe-

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

Table 3 The North Carolina Arboretum Germplasm Repository samples.

ID #	Species	Part	Form	Source	Location	nrDNA	Mix	Hybrid	cpDNA	Mix
NCC1	*A. racemosa*	r/r	powder	NCA	composite 2	*A. racemosa*	no	no	*A. racemosa*	no
NCC2	*A. racemosa*	r/r	powder	NCA	composite 2	*A. racemosa*	no	no	*A. racemosa*	no
NC01a	*A. racemosa*	r/r	powder	NCA	NC					
NC01b	*A. racemosa*	r/r	powder	NCA	NC					
NC01c	*A. racemosa*	r/r	powder	NCA	NC					
NC02a	*A. racemosa*	r/r	powder	NCA	VA					
NC02b	*A. racemosa*	r/r	powder	NCA	VA					
NC02c	*A. racemosa*	r/r	powder	NCA	VA					
NC03a	*A. racemosa*	r/r	powder	NCA	PA					
NC03b	*A. racemosa*	r/r	powder	NCA	PA					
NC03c	*A. racemosa*	r/r	powder	NCA	PA					
NC04a	*A. racemosa*	r/r	powder	NCA	NY					
NC04b	*A. racemosa*	r/r	powder	NCA	NY					
NC04c	*A. racemosa*	r/r	powder	NCA	NY					
NC05a	*A. racemosa*	r/r	powder	NCA	AR					
NC05b	*A. racemosa*	r/r	powder	NCA	AR					
NC05c	*A. racemosa*	r/r	powder	NCA	AR					
NC06a	*A. racemosa*	r/r	powder	NCA	MO					
NC06b	*A. racemosa*	r/r	powder	NCA	MO					
NC06c	*A. racemosa*	r/r	powder	NCA	MO					
NC07a	*A. racemosa*	r/r	powder	NCA	WV					
NC08a	*A. racemosa*	r/r	powder	NCA	WV					
NC08b	*A. racemosa*	r/r	powder	NCA	WV					
NC08c	*A. racemosa*	r/r	powder	NCA	WV					
NC09a	*A. racemosa*	r/r	powder	NCA	WV					
NC09b	*A. racemosa*	r/r	powder	NCA	WV					
NC09c	*A. racemosa*	r/r	powder	NCA	WV					
NC10a	*A. racemosa*	r/r	powder	NCA	VA					
NC11a	*A. racemosa*	r/r	powder	NCA	TN					
NC11b	*A. racemosa*	r/r	powder	NCA	TN					
NC11c	*A. racemosa*	r/r	powder	NCA	TN					
NC12a	*A. racemosa*	r/r	powder	NCA	PA					
NC12b	*A. racemosa*	r/r	powder	NCA	PA					
NC12c	*A. racemosa*	r/r	powder	NCA	PA					
NC13a	*A. racemosa*	r/r	powder	NCA	NY					
NC13b	*A. racemosa*	r/r	powder	NCA	NY					
NC13c	*A. racemosa*	r/r	powder	NCA	NY					
NC14a	*A. racemosa*	r/r	powder	NCA	NC					
NC14b	*A. racemosa*	r/r	powder	NCA	NC					
NC14c	*A. racemosa*	r/r	powder	NCA	NC					
NC15a	*A. racemosa*	r/r	powder	NCA	NC					
NC15b	*A. racemosa*	r/r	powder	NCA	NC					
NC15c	*A. racemosa*	r/r	powder	NCA	NC					
NC16a	*A. racemosa*	r/r	powder	NCA	NC					
NC16b	*A. racemosa*	r/r	powder	NCA	NC					
NC16c	*A. racemosa*	r/r	powder	NCA	NC					
NC18a	*A. racemosa*	r/r	powder	NCA	MD					
NC19a	*A. racemosa*	r/r	powder	NCA	KY					
NC19b	*A. racemosa*	r/r	powder	NCA	KY					
NC20a	*A. racemosa*	r/r	powder	NCA	IN					
NC20b	*A. racemosa*	r/r	powder	NCA	IN					
NC20c	*A. racemosa*	r/r	powder	NCA	IN					
NC21a	*A. racemosa*	r/r	powder	NCA	DE					
NC21b	*A. racemosa*	r/r	powder	NCA	DE					
NC21c	*A. racemosa*	r/r	powder	NCA	DE					
NC22a	*A. racemosa*	r/r	powder	NCA	DE					

nrDNA: nuclear ribosomal DNA; cpDNA: chloroplast DNA; r/r: roots/rhizomes

cies [*Actaea cimicifuga* L., *Actaea pachypoda* Ell., *Actaea podocarpa* DC., and *Actaea rubra* (Aiton) Willd.] purchased from AHP (**Table 1**) were submitted to PCA and SIMCA. The same samples were submitted to PCA and SIMCA so that the FIMS and NMR results may be compared.

Fig. 2 comprises PCA score plots and SIMCA influence plots obtained by FIMS in the left column and NMR in the right column. There are strong similarities between the responses of the two methods as can be seen by the PCA scores. For both instruments: 1) the *A. racemosa* clustered separately from the other species, 2)

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

Table 4 Commercial roots and supplements.

ID #	Species	Part	Form	Source	Location	nrDNA	Mix	Hybrid	cpDNA	Mix
CA01	*A. racemosa*	r/r	r/r	–	Liaoning, China	*A. dahurica*	yes	n/a	*A. dahurica*	no
CA02	*A. heracleifolia*	r/r	r/r	–	Heilongjiang, China	*A. dahurica*	yes	n/a	No ID	yes
CA03	*A. heracleifolia*	r/r	r/r	–	Sichuan, China	Acanthaceae	no	n/a	*Baphicacanthus*	no
CA04	*A. foetida*	r/r	r/r	–	Hebei, China	*Eurotium* sp.	yes	n/a	No ID	n/a
CA05	*A. foetida*	r/r	r/r	–	suzhou,China	*A. dahurica*	yes	n/a	*A. dahurica*	no
CA06	*A. dahurica*	r/r	r/r	–	Hebei, China	No ID	yes	n/a	*A. dahurica*	yes
CA07	*A. dahurica*	r/r	r/r	–	Sichuan, China	*Pichia* sp.	yes	n/a	*A. dahurica*	no
CA08	*A. dahurica*	r/r	r/r	–	North Korea	*A. dahurica*	yes	no	*A. dahurica*	no
CA09	*A. foetida*	r/r	r/r	–	North Korea	*A. brachycarpa*	yes	no	*A. brachycarpa*	no
CA10	*A. foetida*	r/r	r/r	–	Yunnan, China	*A. dahurica*	yes	yes	*A. dahurica*	no
CA11	*A. foetida*	r/r	r/r	–	Henan, China	*A. dahurica*	yes	yes	*A. dahurica*	yes
CA12	*Vernonia aspera*	r/r	r/r	–	Hebei, China	*Eupatorium*	no	n/a	*E. fortunei*	no
CA13	*V. aspera*	r/r	r/r	–	Yunnan, China	*Astereae*	yes	n/a	*E. fortunei*	no
CA14	*V. aspera*	r/r	r/r	–	Yunnan, China					
CS01	*A. racemosa*	–	tab		NA					
CS02	*A. racemosa*	–	tab		NA					
CS03	*A. racemosa*	–	liq		NA					
CS04	*A. racemosa*	–	liq		NA					
CS05	*A. racemosa*	–	liq		NA					
CS06	*A. racemosa*	–	liq		NA					
CS07	*A. racemosa*	–	liq		NA					
CS08	*A. racemosa*	–	cap		NA	*A. racemosa*	no	no	*A. racemosa*	no
CS09	*A. racemosa*	–	cap		NA	*A. racemosa*	no	no	*A. racemosa*	no
CS10	*A. racemosa*	–	cap		?	*A. racemosa*	no	no	*A. racemosa*	no
CS11	*A. racemosa*	–	cap		?	no DNA	n/a	no	no DNA	n/a
CS12	*A. racemosa*	–	cap		?	*A. brachycarpa*	no	no	pqs	n/a
CS13	*A. racemosa*	–	cap		?	*A. racemosa*	yes	no	pqs	n/a
CS14	*A. racemosa*	–	cap		?	*Oryza sativa*	yes	n/a	*O. sativa*	no

nrDNA: nuclear ribosomal DNA; cpDNA: chloroplast DNA; r/r: roots/rhizomes; n/a: not analyzed; pqs: poor quality sequence; NA: not available; tab: tablet; liq: liquid; cap: capsule

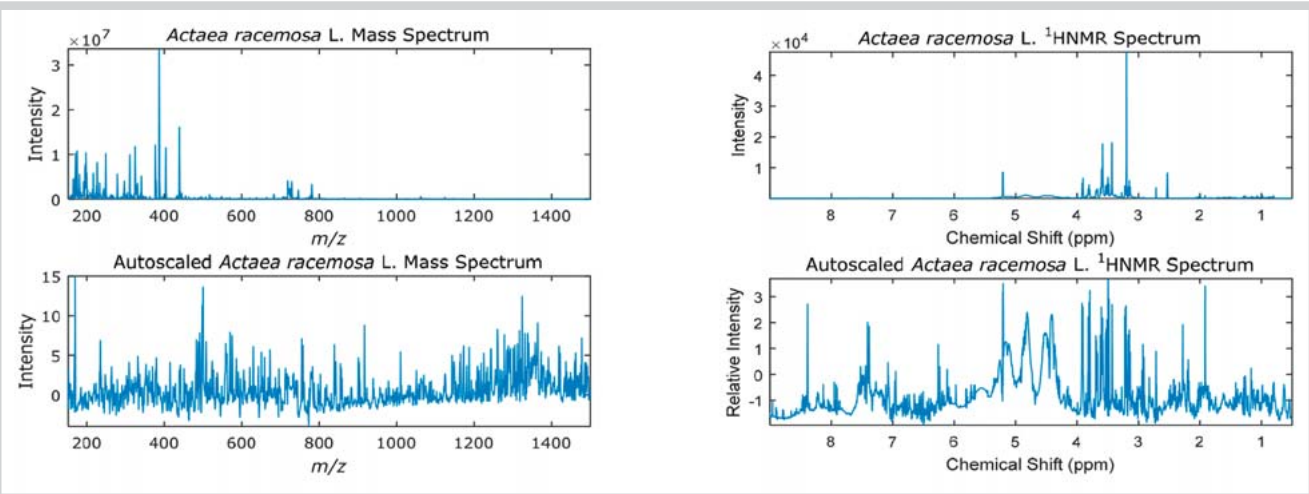

Fig. 1 Comparison of *A. racemosa* spectra before and after preprocessing, which comprised normalization to unit length followed by autoscaling to the *A. racemosa* species. For the NMR, the DMSO solvent peak was removed at 2.54 ppm before preprocessing. (Color figure available online only.)

sample BCR02, identified by AHP as *A. pachypoda*, appears in the *A. racemosa*, cluster, and 3) sample BCR16, identified by AHP as *A. racemosa*, appears outside the *A. racemosa* cluster. Samples BCR02 and BCR16 were analyzed twice by NMR two months apart; the repeat analyses agreed well with the original analyses. In general, the discrimination between the *Actaea* species provided by the two methods is in close agreement.

The influence plots (i.e., Q statistic plotted as a function of the Hotelling T^2 statistic) are given at the bottom of ▶ **Fig. 2** for the FIMS and NMR data, respectively, for single component SIMCA models based on *A. racemosa* identified in ▶ **Table 1**. The T^2 and Q statistics provide the variance accounted for by the SIMCA model and the residual variance unaccounted for by the model, respectively. In both cases, preprocessing consisted of normalization and autoscaling. For both instruments, the models for

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

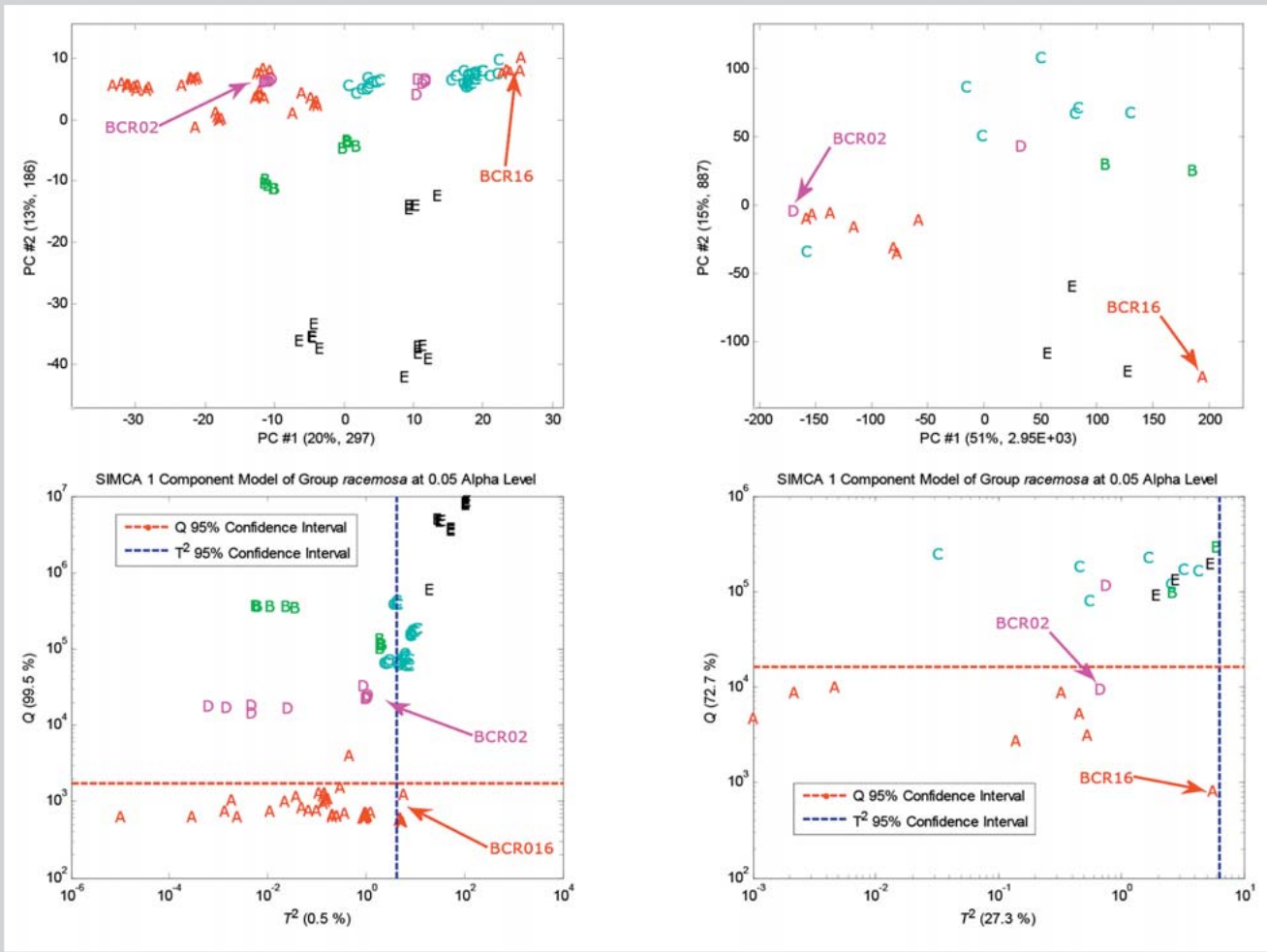

Fig. 2 Top left, scores of the preprocessed mass spectra; bottom left, one component SIMCA influence plot based on the AHP *A. racemosa* spectra. Top right, scores of the preprocessed NMR spectra; bottom right, one component SIMCA influence plot of the same samples at the mass spectra. **A** (red) *A. racemosa*, **B** (green) *A. rubra*, **C** (cyan) *A. cimicifuga*, **D** (magenta) *A. pachypoda*, **E** (black) *A. podocarpa*. (Color figure available online only.)

A. racemosa included sample BCR16, even though it did not cluster with the other *A. racemosa* scores in the PCA plots. The SIMCA models were based on one principal component, and the autoscale mean and standard deviations were only calculated from the set of *A. racemosa* spectra.

It is not appropriate to use influence plots without validation. Models are traditionally validated using bootstrapping. However, the limited number of samples in this phase of the project makes bootstrapping problematic. Dropping a single sample from the NMR data or a set of five repeats for a single sample from the FIMS data can produce dramatically different results. Still, the results for bootstrapping showed sensitivities of 91.4% and 91.7% for FIMS and NMR, respectively (data not shown). The specificities were 100.0% for both methods.

The results in ▶ **Fig. 2** show the close agreement of the models based on the FIMS and NMR spectral fingerprinting and show that both are highly sensitive to changes in the sample chemical composition. Both methods can discriminate between the five *Actaea* species. Thus, for the purpose of identification and authentication, both FIMS and NMR perform equally well.

The third method used to characterize the *Actaea* species was DNA sequencing. Barcodes were determined at two loci: one nuclear ribosomal (nrDNA) and one chloroplast DNA (cpDNA). The nrDNA were examined to determine if there was a mixture of DNA (contamination) or if the DNA represented a hybrid (species interbreeding). The cpDNA was also used to look for mixtures. Results for the DNA sequencing are presented in ▶ **Tables 1–4**.

In general, DNA sequencing confirmed the species identification provided by AHP (▶ **Table 1**). The exceptions were samples BCR03 and BCR18–BCR24. In each case, identification was complicated by the purity of the material. BCR03 was identified as *A. pachypoda* using nrDNA, but gave a poor quality sequence (PQS) with cpDNA. Samples BCR18 to BCR21 appeared moldy upon visual examination and were identified as the fungus *Erotium* by nrDNA. Both loci (nrDNA and cpDNA) indicated a mixture of DNA materials. BCR21 to BCR24 were identified as *Actaea dahurica* (Turcz. ex Fisch. & C. A.Mey.) Franch. and not *A. cimicifuga*. Previously, next generation sequencing of SRM 3295 *A. racemosa* Rhizome had revealed that 2% of the DNA present was fungal [29]. *Eurotium herbariorum*, for example, was found in SRM 3295 and is a soil born fungus that is frequently found on dried plant products and is common in stored seeds. Thus, identification of BCR18 to BCR21 as the *Eurotium* genus is not surprising. These results do not preclude that *A. cimicifuga* is present in samples BCR18 to BCR21. The other samples in this study were not analyzed using next generation sequencing, which is unfortunate

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

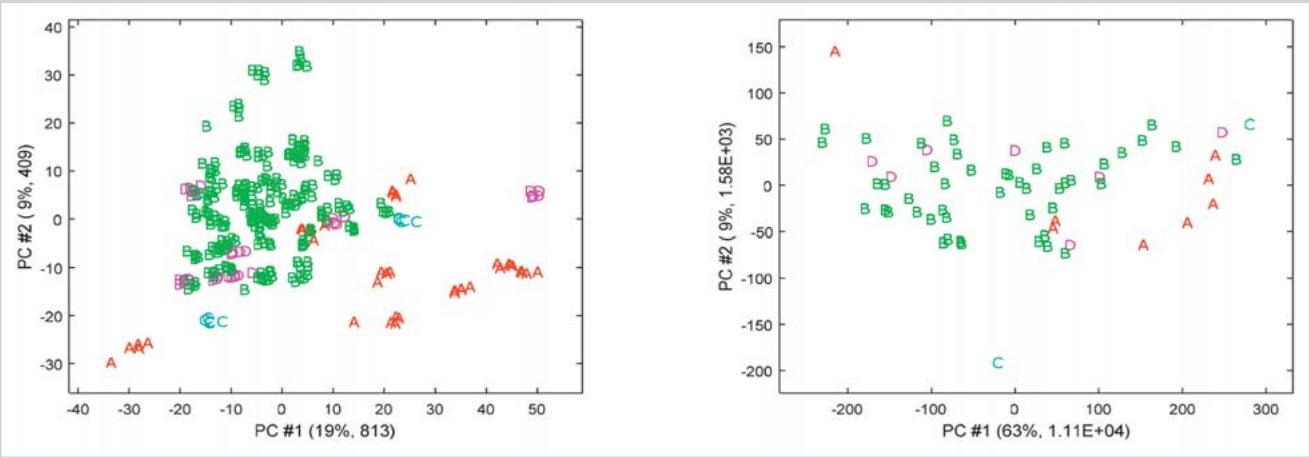

Fig. 3 Left, PCA scores of preprocessed *A. racemosa* mass spectra with respect to the different suppliers; right, PCA scores of preprocessed *A. racemosa* NMR spectra; right bottom, corresponding influence plot for a one component SIMCA model of the NMR spectra. **A** (red) American Herbal Pharmacopoeia, **B** (green) North Carolina Arboretum, **C** (cyan) National Institutes of Standards, **D** (magenta) Strategic Sources. (Color figure available online only.)

because understanding the level of fungal load may be a significant factor in interpreting the chemical profiles.

Interestingly, samples BCR02 and BCR16, which appeared to be in the wrong clusters in **Fig. 2**, were in fact accurately identified as *A. pachypoda* and *A. racemosa*, respectively, by DNA sequencing. The PCA scores in **Fig. 2** suggested that they were misidentified and led to the purchase of new samples from AHP. The new samples were analyzed by MS, NMR, and DNA sequencing, and produced the same results as initially obtained. The positions of BCR02 and BCR16 suggest that they contain components (or are missing components) that make them slightly different from the other samples of the same species. However, for both FIMS and NMR, the influence plots showed that the spectral fingerprints of BCR02 samples were closest to the other *A. pachypoda* samples, and fingerprints for BCR16 were closest to *A. racemosa*. These results emphasize the fact that these are complex biological samples and the phenotypes display far greater variation than the genotype.

As listed in **Tables 1–3**, authentic *A. racemosa* roots/rhizomes were obtained from four major sources. **Fig. 3** comprises PCA score plots from spectra acquired by FIMS and NMR for *A. racemosa* samples from the American Herbal Pharmacopoeia (AHP), The North Carolina Arboretum Germplasm Repository (TNCAGR), Strategic Sources, Inc. (SSI), and the National Institute of Standards and Technology (SRM 3295, Black Cohosh Rhizome). TNCAGR furnished two composite *A. racemosa* materials that were used for method development and 55 samples collected from 22 sites from the east coast and as far west as Missouri.

The scores in the PCA plot in **Fig. 3** characterize the variation between the sources of the authentic *A. racemosa* samples. For MS, the differentiation of the sources is more apparent when incorporating the third principal component and viewed in three dimensions (score plot not shown). However, the distinction between sources was seen clearly by the NMR scores for the first two components. The samples were prepared independently for FIMS and run on the same day in a random order. When the FIMS and NMR are subjected to class modeling, the influence plots (using *A. racemosa* samples from AHP, TNCAGR, or SS as the class model) show specificities (fraction of the non-model samples identified as non-model) ranging from 75% to 99% (plots not shown). This indicates that the samples from the three sources are different from each other. Similarly, pooled ANOVA showed that the probability that the means of the three sources were the same was < 0.0001 (data not shown).

The samples were prepared independently for FIMS and run on the same day in a random order. The same samples were dried, reconstituted in DMSO, and analyzed randomly by NMR. The agreement of the statistical analysis of the data from the two methods strongly suggests chemical differences between the different sources of authentic samples. These differences may arise from the geographic source of the samples (e.g., local genetic variations, different growth conditions, and/or different local pathogens) and/or systematic differences in post-harvest processing.

Metadata for the *A. racemosa* roots are presented in **Tables 1–3**. Samples from TNCAGR and SSI were collected in the U.S. TNCAGR samples were originally harvested from native populations six years ago and cultivated under controlled conditions. All samples were harvested in October 2012. In general, most of the SS samples come from further west. The AHP samples come from a variety of sources across the U.S. as well as from Canada, China, and commercial sources. NIST SRM 3295 was obtained in bulk from a commercial source and the point of origin has not been disclosed. In most cases, the year of harvest is not available. DNA sequencing confirmed that the samples identified as *A. racemosa* by their supplier were genetically consistent with *A. racemosa*. The misidentification of one sample (BCR03) was discussed above. This sample was removed from the analyses.

Added insight into the variation within species can be gained by taking a closer look at the *A. racemosa* samples furnished by TNCAGR. **Fig. 4** contains PCA score plots for those ten sites for which three samples were received (only one or two samples were received for the other five sites) and were analyzed by both FIMS and NMR. Both sets of profiles have a non-homogeneous distribution of the data from the different sites. The FIMS data demonstrate that the within site variance is considerably less than the between site variance and data from both methods demonstrate that the variance between sites is considerably less than the total variance for all sites. This was verified using pooled ANOVA. The probability that the means of the ten sites were sim-

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

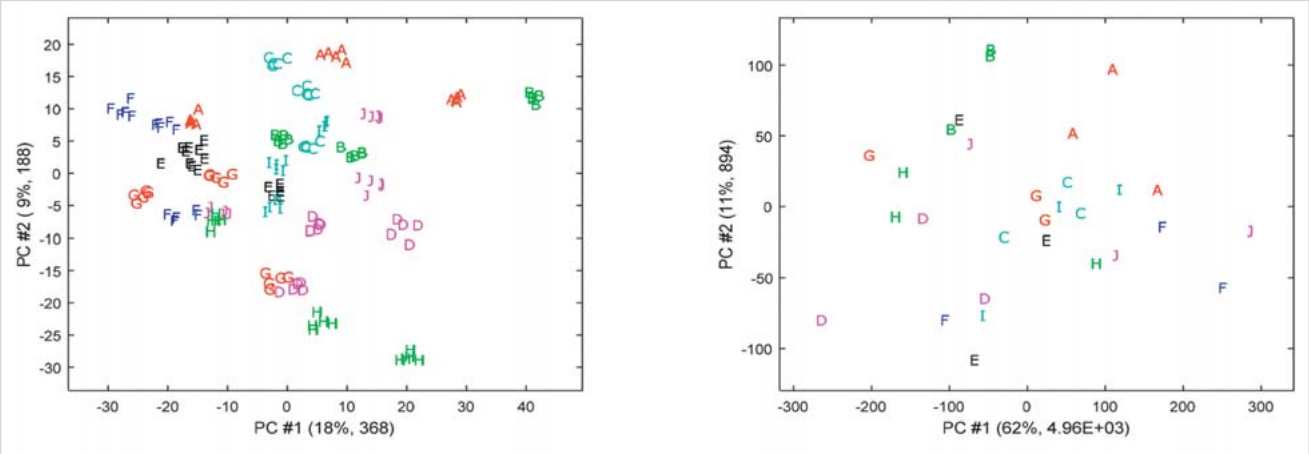

Fig. 4 Left, PCA scores of the preprocessed mass spectra with respect to 10 sites from the North Carolina Arboretum; right, PCA scores of the preprocessed NMR spectra with respect to the same 10 sites. (Each site has a different letter and color.) (Color figure available online only.)

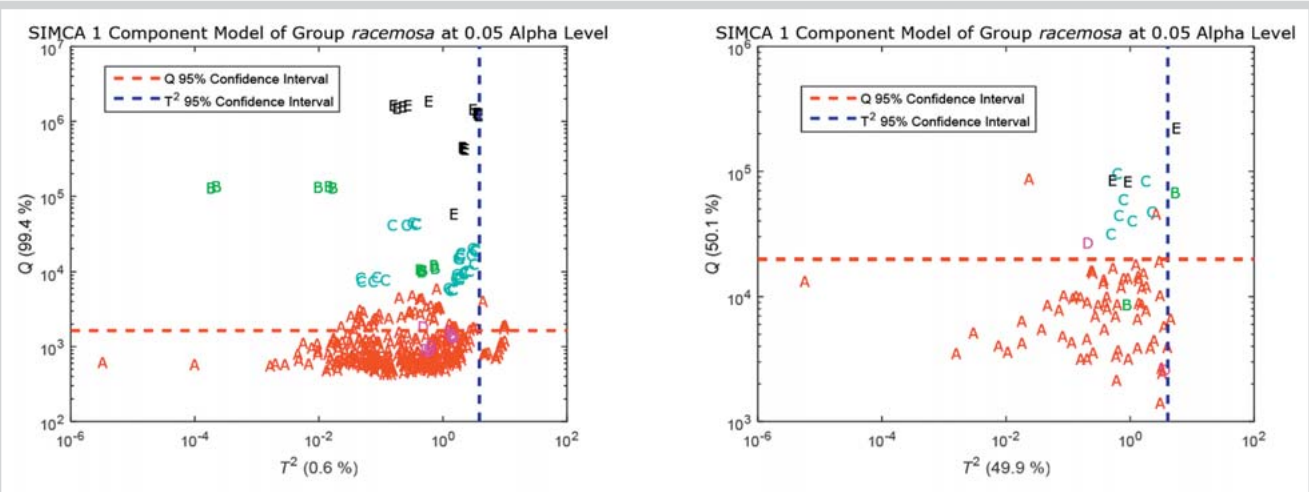

Fig. 5 Influence plots of the preprocessed mass (left) and NMR (right) spectra for the *Actaea* roots of the five species and all suppliers. **A** (red) *A. racemosa*, **B** (green) *A. rubra*, **C** (cyan) *A. cimicifuga*, **D** (magenta) *A. pachypoda*, **E** (black) *A. podocarpa*. (Color figure available online only.)

ilar was < 0.0001 (data not shown). Thus, the individual sites provide subclusters within the *A. racemosa* cluster.

Differences in the chemical fingerprints of the *A. racemosa* samples from the 22 sites could arise from a number of sources. Environmental conditions (e.g., temperature, sunlight, rainfall, soil quality, and altitude) could be important factors. Post-harvest handling (e.g., drying conditions, shelf life, and storage temperatures) can influence enzymatic changes and long-term composition. Isolated plant colonies may also be subject to local genetic mutations. It has also been hypothesized that some of the purported health promoting chemical components of *A. racemosa* may come from endophytic fungi and not the plant itself [30]. Genetic mutation and endophytic fungi variation could potentially be major causes of variation between TNCAGR sites as all samples were grown under separate but similar conditions since their harvest from their original site six years ago. Future studies using second generation sequencing to determine the level of endophytic fungi will be informative.

Fig. 5 presents the influence maps for a single component SIMCA model based on all of the *A. racemosa* samples using the FIMS and NMR spectra. **Fig. 5** is similar to **Fig. 2** influence plots, except the spectra from *A. racemosa* samples from all four sources are used for the class model to test the other four *Actaea* species obtained from AHP. The Q statistic is effective for differentiating *A. racemosa* from the other species with the exception of *A. pachypoda* for FIMS and *A. rubra* for NMR. There are a number of *A. racemosa* samples that fall above the confidence boundary.

As stated earlier, it is not appropriate to use influence plots without validation. Models are traditionally validated using independent sets of data. In this study, the *A. racemosa* fingerprints from 65 samples comprise the model building set of data and the 14 samples from the other species correspond to the negative spectra to evaluate the specificity of the model. To validate the model, the 65 *A. racemosa* samples were randomly split into quarters so that 75% were used for constructing the model and the other 25% for evaluating the model. The FIMS data were randomly parti-

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

	Modeling ($\alpha = 0.05$)			Classification	
Data set	Number	FOAM	SIMCA	FuRES	PLS-DA
MS sensitivity	324	96.7 ± 0.2%	91.1 ± 0.2%	98.2 ± 0.3%	95.4 ± 0.3%
MS specificity	70	86.7 ± 0.1%	85.7 ± 0.1%	59.4 ± 0.8%	63.4 ± 0.3%
NMR sensitivity	65	91.0 ± 0.3%	77.7 ± 0.3%	96.2 ± 0.3%	98.4 ± 0.1%
NMR specificity	14	78.5 ± 0.6%	85.7 ± 0.1%	68.6 ± 0.3%	66.7 ± 0.4%

Table 5 Modeling and classification results with 100 × 4 bootstrapped Latin partitions.

tioned by samples, i.e., in multiples of five spectra. This procedure is comparable to the partitioning of the NMR data, which only consisted of a single spectrum per sample. For modeling, the sensitivity was determined as the fraction of *A. racemosa* samples that were correctly identified as *A. racemosa*. The specificity is the fraction of the other species that were correctly identified as not belonging to the *A. racemosa* model.

Table 5 reports the validation results obtained using both class modeling and classification. For class modeling, SIMCA and FOAM (see Materials and Methods section) were compared using 100 bootstraps with four Latin partitions. Neither SIMCA nor FOAM were optimized in any way, but were used with standard parameters for all the evaluations with the exception that FOAM allowed the internal bootstrap validation to partition by sample to provide a better estimate of the residual error. While the same modification could have been made to SIMCA, it was left unmodified so that it would remain consistent with SIMCA calculations that are commercially available. SIMCA and FOAM tended to perform complementarily with one furnishing higher specificity while the other yielded higher sensitivity.

FIMS had the best performance with a 96.7 ± 0.2% average sensitivity and an average specificity of 86.7 ± 0.1% using FOAM. The SIMCA results were slightly worse with 91.1 ± 0.2% and 85.7 ± 0.1% average sensitivity and specificity, respectively. For NMR, the FOAM average sensitivity and specificity were 91.0 ± 0.3% and 78.5 ± 0.6%, respectively. For SIMCA, the average specificity of 85.7 ± 0.1% was greater than the average sensitivity of 77.7 ± 0.3%. NMR was at a slight disadvantage because fewer spectra were available for building statistical models. These results for the specificities are not so bad when one considers that the *A. rubra* and *A. pachypoda* species only contained two samples each and neither the FOAM nor the SIMCA models were optimized in any way.

The relationship between sensitivity and specificity is dependent on the confidence limit or criterion for group membership that can be controlled to favor the sensitivity or specificity of the model. The 95% confidence interval, which is the red horizontal dashed lines in **Fig. 5 A, B**, is an example of the model acceptance criterion. A useful approach to evaluate a model is the receiver operating characteristic (ROC) curve, which plots the sensitivity with respect to the false positive rate (i.e., 1-specificity) as a function of the models acceptance criterion [31, 32].

The bootstrap Latin partition results were saved to generate the average ROC curves in **Fig. 6**. The best compromise of the sensitivity-specificity tradeoff occurs at the point along the curve closest to the upper left hand corner. By visual inspection one can see that this point corresponds to 90% sensitivity and 90% specificity. In general, FOAM appears to perform a little better that SIMCA. It is not possible to directly compare MS and NMR because much fewer samples were analyzed by NMR.

Two classification methods, PLS-DA and FuRES, were compared. As with the modeling evaluation, the FIMS spectra for *A. racemosa* were partitioned by sample (multiples of five spectra) to be comparable to the partitioning of the NMR spectra. The average classification results of the bootstrapped Latin partitions are reported in **Table 5**. As mentioned in the introduction, classification is a targeted analysis that, in this case, assigns spectra or fingerprints to one of four classes (*A. cimicifuga, A. podocarpa, A. racemosa,* and *A. rubra)*. Specificity, defined as the correct identification for species other than *Racemosa*, is challenging because two of these species had only two samples, and the third species only had three samples. Therefore, models were built from one sample and applied to recognize the other samples of the same species for three of the classes, *A. pachypoda, A. rubra*, and *A. podocarpa.* Average classification results in **Table 5** indicate that PLS-DA and FuRES are comparable for both FIMS and NMR. In this case, when one classifier yielded a greater sensitivity, the other classifier gave a greater specificity. The sensitivities were all above 95% for FIMS and NMR. The specificities were lower and ranged between 60–70%, which was not surprising with the small sample sizes for the non-*A. racemosa* spectra. Classification is most effective if the populations of each class are equal. Despite the excellent sensitivity, the two smallest classes, *A. pachypoda* and *A. rubra,* were misclassified (data not shown). These data emphasize the importance of having large representative sets of samples for classifier construction.

Fig. 7 presents the influence models for FIMS and NMR based on the SIMCA for the same *A. racemosa* samples used in **Fig. 5** to test commercial Asian root samples (**Table 4**, CA01 to CA14). The samples were purported to be *A. racemosa, A. heracleifolia, A. foetida*, and *A. dahurica*. In **Fig. 6**, all of the commercial sample scores are above the 95% Q confidence limit, even the sample of *A. racemosa*. All the samples were subjected to DNA sequencing. Seven of the samples were identified as *A. dahurica*, including the sample purported to be *A. racemosa*. DNA sequencing identified one of the samples as *A. brachycarpa* and four of the samples were other plant materials. All of the samples were a mixture of materials and two could only be identified as a yeast (*Pichia* spec.) and a fungi (*Erotium* spec.).

Fig. 8 again uses the same influence SIMCA plots for FIMS and NMR as **Fig. 5** to test 14 commercial supplements of *A. racemosa* purchased from local stores (**Table 4**, CS01 to CS14). Two were tablets, seven were capsules, and five were in liquid form. All the supplements, with the exception of one capsule (by both FIMS and NMR), were differentiated from *A. racemosa* root at the 95% Q confidence level. Most likely, this capsule (CS09) contained powdered *A. racemos*a root as DNA sequencing confirmed the presence of *A. racemosa* DNA.

A SIMCA model was constructed based on the NIST extracted SRMs (3297 and 3298). Since the SRM materials were obtained from a commercial source that used standard methods for their preparation, it was hypothesized that they would serve as a suitable model for the commercial samples. However, all the commercial supplements were excluded from the model at the 95% confidence level (data not shown). This result indicates that the different preparations of the SRMs and the commercial supplements have resulted in different chemical fingerprints. This result

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

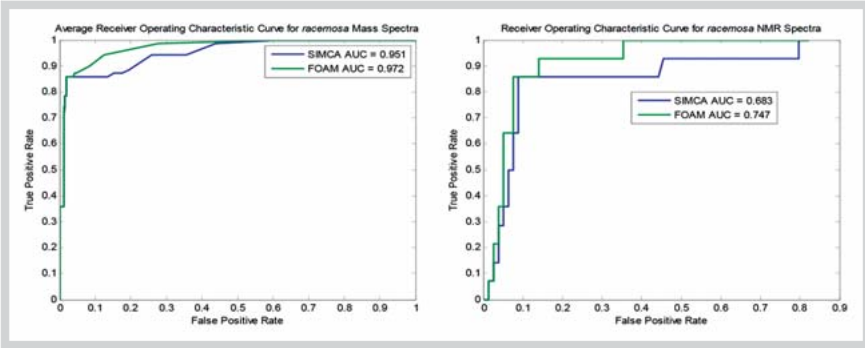

Fig. 6 Average receiver operating curves for 100 × 4 bootstrapped Latin partitions that averaged 100 sensitivities and 400 specificities for the MS and NMR modeling evaluations that were built with the racemosa *A. racemosa* data. (Color figure available online only.)

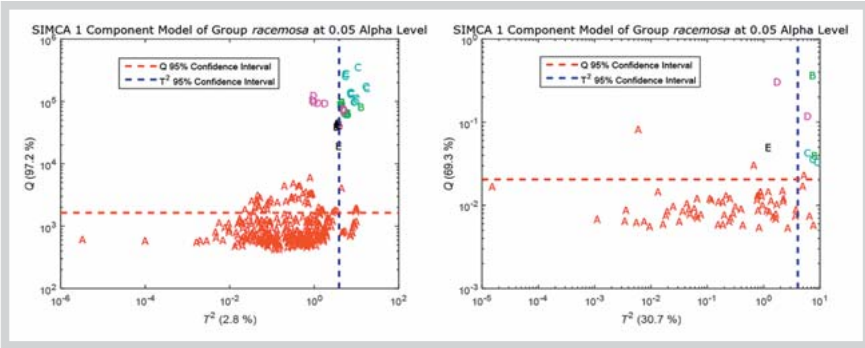

Fig. 7 Influence plots of commercial species on a model built from the *A. racemosa* one component SIMCA reference model. Mass spectra (left) and NMR spectra (right). **A** (red) *A. racemosa*, **B** (green) *A. cimicifuga*, **C** (cyan) *A. dahurica*, **D** (magenta) *A. heracleifolia*, **E** (black) commercial *A. racemosa*. (Color figure available online only.)

does not preclude the presence of *A. racemosa* extract in the supplements.

DNA sequencing showed that four of the seven capsules contained DNA from *A. racemosa*. One contained *A. brachycarpa* DNA and two contained no *Actaea* DNA. Of the latter, one had no DNA at all and the other had only rice DNA (*Oryza sativa* L.; Poaceae), presumably from a rice excipient. Five of the seven showed no indications of a mixture, indicating that the excipient was a refined chemical component, such as a crystalline starch, with no DNA present. None of the liquid supplements were subjected to DNA sequencing as it was assumed there would be no DNA present. These data suggest that a more rigorous metabolomics study is necessary to verify the presence of specific *A. racemosa* marker compounds in the supplements.

In summary, FIMS and NMR produced metabolic fingerprints that were equally capable of discriminating between *Actaea* species and authenticating *A. racemosa* when looking at raw materials. DNA sequencing of the root materials could be used to identify species and was useful in validating anomalies in the clusters observed by both metabolic fingerprinting methods. FIMS and NMR detected differences between sources of samples of *A. racemosa*, suggesting systematic geographical and/or processing differences. All three methods demonstrated that none of the tested commercially available root samples were *A. racemosa*. FIMS and NMR fingerprints indicated that commercially available supplements were not similar to any of the raw *Actaea* materials, suggesting that preparation produced significantly different metabolic profiles. DNA sequencing was successful for identifying four of the seven supplements tested as *A. racemosa*. The others were a different species or had no DNA present. Receiver operating characteristic curves are useful for evaluating and optimizing authentication models of complex data sets.

Materials and Methods

Actaea samples

Root materials from various species of *Actaea*, some vouchered, were collected from four reliable sources and from a variety of commercial sources (**Tables 1–4**). The sources of samples were the American Herbal Pharmacopoeia (AHP; samples BCR01 to BCR24 in **Table 1**), Strategic Sourcing, Inc. (SSI; samples SS01 to SS07 in **Table 2**), The North Carolina Arboretum Germplasm Repository (TNCAGR; samples NCC1, NCC2, and NC01 to NC22 in **Table 3**), and the National Institutes of Standards and Technology (NIST; samples SRM 3295, 3296, 3297, and 3298 in **Table 2**). Commercial root samples were purchased from the Internet and local stores in China (samples CA01 to CA14 in **Table 4**). Commercial liquid, tablet, and capsule supplements were purchased from local stores in Maryland (CS01 to CS14). TNCAGR samples were collected from the permanent national *A. racemosa* germplasm collection in collaboration with the USDA NPGS (National Plant Germplasm System).

DNA sequencing

All root materials and finished products were authenticated using validated DNA sequencing authentication methods for *Actaea* [29] at AuthenTechnologies LLC. The methods utilized consisted of the extraction of total genomic DNA using a modified silica-spin column approach (Qiagen, Inc.). Next, the nuclear (ITS) and chloroplast (psbA-trnH) genes validated by the National Institute of Standards and Technology [29] for black cohosh identification were amplified using a polymerase chain reaction (PCR) machine using standard cycling parameters. The PCR products were visualized using an E-Gel (Invitrogen) apparatus and visualized using a blue LED box. Positive PCR products were then sequenced on a Sanger sequencing (capillary electrophoresis) machine (Applied Biosystems). The resultant sequence from both the forward and

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

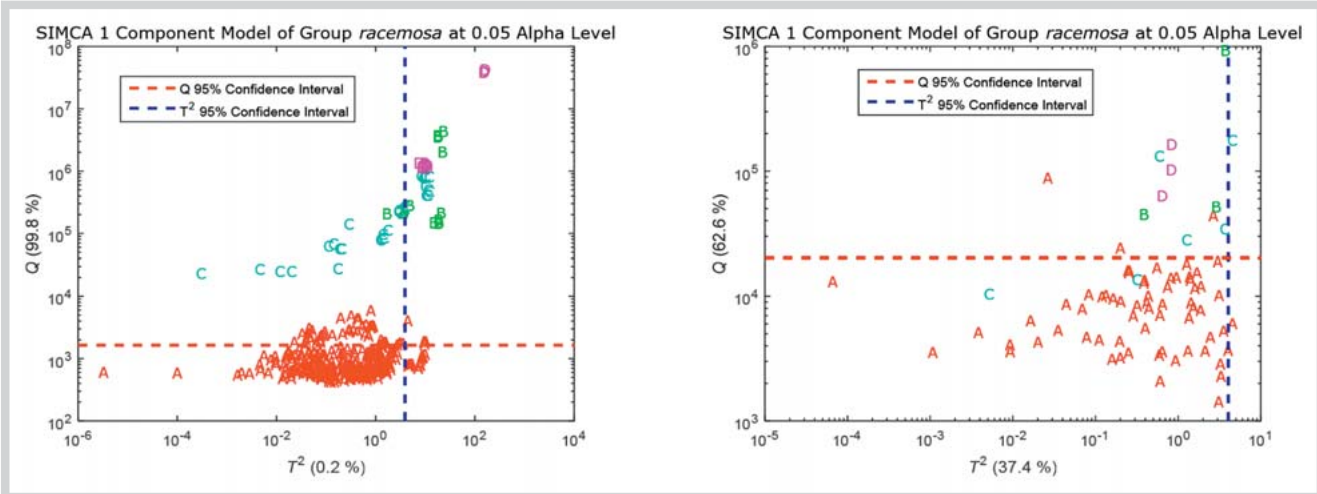

Fig. 8 Influence plots of the *A. racemosa* SIMCA model with respect to different supplement forms. **A** (red) root, **B** (green) tablet, **C** (cyan) capsule, and **D** (magenta) liquid. (Color figure available online only.)

reverse directions was assembled into contigs. Overlapping bases indicative of hybrids were scored using standard IUPAC codes. The final assembled sequences were then aligned in a matrix by eye and compared to reference sequences obtained from authenticated herbarium vouchers for the target and closely related nontarget species of *Actaea*.

Mass spectrometry sample preparation

Root samples were ground into fine powders. Ten mg of each sample were mixed with 5 mL of methanol-water (70–30, v/v) in 15 mL centrifuge tubes and sonicated for 60 min at room temperature. The extracted samples were centrifuged at 5000 × *g* for 10 min (IEC Clinical Centrifuge, Danon/IEC Division). The supernatant was diluted 1 to 10 (v/v) with methanol and filtered through a 17-mm (0.45 µm) PVDF syringe filter (VWR) prior to injection. To avoid errors arising from unexpected degradation of some compounds, the sample analysis was completed within 24 h of the extraction. Tablets were prepared the same as root samples. Capsules were opened and the contents were emptied onto a weighing paper. Ten mg were mixed with 5 mL of methanol-water (70–30, v/v) in a 15-mL centrifuge tube and then treated in the same manner as the root samples. Ten µL of liquid supplement were mixed with 5 ml of methanol-water (70–30, v/v) in a 15-mL centrifuge tube and then treated in the same manner as the root samples.

Nuclear magnetic resonance sample preparation

Aliquots of the samples prepared in methanol-water for analysis by FIMS were taken to dryness for transportation to the Bruker Bio-Spin laboratory. Samples were prepared for NMR by dissolving 25 mg of each dried sample in 1.0 mL of DMSO-d_6 containing 0.47 mM DSS to provide a final concentration of 25 mg/ml. Each sample was vortexed for 1 min and sonicated for 5 min. The samples were then centrifuged for 15 min at 13 500 rpm (Eppendorf 5810 R, Eppendorf AG) to remove any undissolved materials. Then, 600 µL of the supernatant was transferred to a 5-mm NMR tube (Wilmad PP-5 and Bruker Z107374) for spectroscopy.

Mass spectrometry instrumentation

The FIMS system consisted of a Q Exactive mass spectrometer (Thermo Fisher Scientific) with an Agilent 1200 HPLC system (a quaternary pump with a vacuum degasser, a thermostated column compartment, an autosampler, and a diode array detector). The flow injection used a guard column (Adsorbosphere AllGuard Cartridge, C18, 5 µm, 4.6 × 7.5 mm, Alltech Associates, Inc.) to minimize potential contamination of the FIMS system. Mobile phases consisted of 0.1% formic acid in H_2O (A) and 0.1% formic acid in acetonitrile (B) with isocratic elution at 50 : 50 (v/v) at a flow rate of 0.5 mL/min for 2 min. Electrospray ionization was performed in the negative ion mode from *m/z* 150–1500 to obtain the FIMS fingerprints. The following conditions were used for the mass spectrometer: sheath gas flow rate, 80 (arbitrary units); aux gas flow rate, 10 (arbitrary units); spray voltage, 4.50 kV; heated capillary temperature, 220 °C; capillary voltage, 4.0 V; tube lens offset, 25 V. The injection volume for all samples was 10 µL.

Nuclear magnetic resonance instrumentation

All NMR experiments were performed with a Bruker AVANCE III spectrometer (600.13 MHz) at 298 K with a Bruker 5-mm TCI CryoProbe. Spectra were collected using the software program TopSpin 3.1 (Bruker BioSpin). One-dimensional proton nuclear Overhauser effect spectroscopy with an inverse gated decoupling pulse sequence (noesyigld1 d) using a base opt filter was performed using 64 scans, four dummy scans, and 65 536 data points. A relaxation delay of 10 s with a mixing time of 0.01 seconds was used to allow an acquisition time of 14 min for each experiment.

Sample analysis

The sequence of the samples was randomized for both FIMS and NMR. For FIMS, each sample in ◉ **Tables 1–4** (with the exception of BCR21 which was not run) was run five times for a total of 1140 analyses. After running each sample once in random order, a new random sequence of measurements was made. Spectra were summed over the 1.0 min interval from 0.5 min to 1.5 min of the sample bolus.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

For NMR, each sample was run only once. Exceptions were BCR03 and BCR21, which were not run, and BCR02 and BCR16, which were analyzed twice. Each of the second analyses was on a separate day two weeks later to check for possible mislabeling of the samples. This sampling provided a total of 130 analyses. The NMR spectra were acquired from − 3.0 to 16.0 ppm, but only the region from 0.5 to 9.0 ppm was used in this study.

Mass spectrometry data processing

The FIMS fingerprints of each sample were mass spectra, i.e., ion counts with respect to the mass-to-charge ratio for a range of *m/z* 150 to 1500. The spectra were exported as Excel files (Microsoft, Inc.) for data preprocessing and then imported into either Solo (Eigenvector Research, Inc.), for principal component analysis (PCA) and soft independent modeling of class analogy (SIMCA), or to MATLAB 2014a (MathWorks) for classification and validation studies.

Nuclear magnetic resonance data processing

The NMR fingerprints for each sample were spectra, i.e., signal intensity with respect to chemical shift from 0.5 ppm to 9.0 ppm. The spectra were exported as Excel files and then imported into Solo or MATLAB 2014a. Prior to preprocessing, the solvent peaks were removed at 2.53 ppm (DMSO) and 3.19 ppm (methanol) by excising the respective ranges 2.51–2.56 ppm and 3.18–3.21 ppm from all of the spectra.

Chemometrics

The same processing was used for the FIMS and NMR data. The spectra were normalized to unit vector length (i.e., the sum of the squares of the data points for each spectrum was unity), autoscaled, and mean-centered prior to PCA or SIMCA.

Classifiers

There are two general types of classifiers: modeling (soft modeling or one-class classifiers) and classification (hard modeling or multiclass classifiers). The first category comprised SIMCA [22] and FOAM [23] and the latter comprised PLS-DA [24] and FuRES [25]. All four approaches (SIMCA, FOAM, PLS-DA, and FuRES) were applied to the data in this study.

SIMCA models were constrained to a single component. FOAM was used with standard parameters for the fuzzy grid encoding. The grid size was 100 and a 19-point triangular fuzzy membership function was applied to the intensities of the FIMS and NMR spectra. Both SIMCA and FOAM used the Q statistic to determine the fit to the model at a 95% confidence interval. A simple empirical method set the boundary condition to distance that would exclude 5% of the objects in the calibration set. FOAM, however, uses an internal bootstrap Latin partition to determine the average residual error for the objects in the calibration set. The boundary condition was calculated from the average residual errors. FOAM was later modified to estimate the error by partitioning by sample as opposed to spectrum, because the internal bootstrap was underestimating the error. A similar approach could be used with SIMCA and the results would be similar. Note that the internal bootstrap error estimation is only applied to the training set of data.

A self-optimizing version of PLS-DA was used. This method applies a bootstrap Latin partition [33] to the calibration data set and determines the number of components (i.e., latent variables) that minimizes the average prediction error across the bootstraps. For all cases, the calibration data was divided into two partitions and averaged ten times within the PLS-DA computation. A model is constructed using the entire calibration data with the number of components that achieved the lowest prediction error. There are no adjustable parameters for FuRES, the softness of the model is determined by maximizing the magnitude of the first derivative of the fuzzy entropy of classification with respect to the computational temperature (i.e., discriminant vector length).

Validation

The NMR data contained two samples for which repeat measurements were made. For the sensitivity and specificity calculations, the replicate samples that were collected last were removed from the evaluation, so that each sample had only a single spectrum in the data set. The FIMS data set had 81 samples while the NMR data set had 77 samples. The FIMS data set was reduced to the same set of 77 samples that were measured by NMR. In addition, both data sets were reduced to subsets of 20 samples that corresponded to samples obtained from AHP. Validation was accomplished by partitioning the data by sample, so that all five replicates of the same sample would be in either the calibration set or prediction set. This mode allows the FIMS results to be compared with the NMR results. In addition, this method is more rigorous and assesses the performance to generalize to new samples, which is most typically encountered in practical applications.

Prior to classification the principal component transform (PCT) [34], which is a form of lossless compression, was applied to the data to increase the speed for FuRES and PLS-DA. For each bootstrap, the principal components were constructed from the calibration data, and then the prediction data were projected onto those same components. The PCT was unnecessary for SIMCA and FOAM because both of these methods are computationally very fast.

Pooled analysis of variance

This method has been previously described [10]. In brief, classic ANOVA is performed for every variable in the MS or NMR spectra and the resulting ratios are pooled to obtain an overall F value.

Acknowledgements

▼

This research is supported by the Agricultural Research Service of the U.S. Department of Agriculture and an Interagency Agreement with the Office of Dietary Supplements of the National Institutes of Health.

Conflict of Interest

▼

The authors declare no conflicts of interest.

Affiliations

[1] Food Composition and Methods Development Laboratory, Beltsville Human Nutrition Research Center, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, MD, USA
[2] Bruker BioSpin, Billerica, MA, USA
[3] The North Carolina Arboretum Germplasm Repository, Asheville, NC, USA
[4] AuthenTechnologies LLC, Richmond, CA, USA
[5] Center for Intelligent Chemical Instrumentation, Department of Chemistry and Biochemistry, Clippinger Laboratories, Ohio University, Athens, OH, USA
[6] Strategic Sourcing, Inc., Banner Elk, NC, USA

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

References

1 *Blumenthal M.* Herb sales down 7.4 percent in mainstream market; garlic is top-selling herb; herb combinations see increase. HerbalGram 2005; 66: 63

2 *Hildreth J, Hrabeta-Robinson E, Applequist W, Betz L.* Standard operating procedure for the collection and preparation of voucher plant specimens for use in the nutraceutical industry. Anal Bioanal Chem 2007; 389: 13–17

3 *Hall RD.* Plant metabolomics from holistic hope, to hype, to hot topic. New Phytol 2006; 169: 453–468

4 *AOAC International; Guideline Working Group.* AOAC INTERNATIONAL guidelines for validation of botanical identification methods. J AOAC Int 2012; 95: 268–272

5 *LaBudde R, Harnly JM.* Probability of identification (POI): a statistical model for the validation of qualitative botanical identification methods. J AOAC Int 2012; 95: 273–285

6 *Ma C, Kavalier AR, Jiang B, Kennelly EJ.* Metabolic profiling of *Actaea* species extracts using high performance liquid chromatography coupled with electrospray ionization time-of-flight mass spectrometry. J Chromatogr A 2011; 1218: 1461–1476

7 *Qiu F, McAlpine JB, Lankin DC, Burton I, Karakach T, Chen SN, Pauli GF.* 2D NMR barcoding and differential analysis of complex mixtures for chemical identification: the *Actaea* terpenes. Anal Chem 2014; 86: 3964–3972

8 *Masada-Atsumi S, Kumeta Y, Takahashi Y, Hakamatsuka T, Goda Y.* Evaluation of the botanical origin of black cohosh products by generic and chemical analyses. Biol Pharm Bull 2014; 37: 454–460

9 *He K, Pauli GF, Zheng B, Wang H, Bai N, Peng T, Roller M, Zheng Q. Cimicifuga* species identification by high performance liquid chromatography-photodiode array/mass spectrometric/evaporative light scattering detection for quality control of black cohosh products. J Chromatogr A 2006; 1112: 241–254

10 *Harnly JM, Harrington PB.* Adulteration of American with Asian ginseng: spectral addition and experimental verification of probability of identification. J AOAC Int 2013; 96: 1258–1265

11 *Harnly JM, Luthria DL, Chen P.* Detection of adulterated *Ginkgo biloba* supplements using chromatographic and spectral fingerprints. J AOAC Int 2012; 95: 1579–1587

12 *Harnly JM, Mukhopadhyay S, Lin LZ, Luthria DL.* A comparison of analytical and data preprocessing methods for spectral fingerprinting. Appl Spectrosc 2011; 65: 250–259

13 *Luthria DL, Lin LZ, Robbins RJ, Finley JW, Banuelos GS, Harnly JM.* Discriminating between cultivars and treatments of broccoli using mass spectral fingerprinting and analysis of variance principal component analysis. J Agric Food Chem 2008; 56: 9819–9827

14 *Chen P, Luthria D, Harrington PB, Harnly JM.* Discrimination among *Panax* species using spectral fingerprinting. J AOAC Int 2011; 94: 1411–1421

15 *Huang H, Sun J, McCoy JA, Zhong H, Fletcher EJ, Harnly JM, Chen P.* Use of flow injection mass spectrometric fingerprinting (FIMS) and chemometrics for differentiation of three black cohosh species. Spectrochim Acta 2014, in press

16 *Chen P, Sun J, Ford P.* Differentiation of the four major species of cinnamons (*C. burmannii, C. verum, C. cassia*, and *C. loureiroi*) using a flow injection mass spectrometric (FIMS) fingerprinting method. J Agric Food Chem 2014; 62: 2516–2521

17 *Safer S, Cicek A, Pieri V, Schwaiger P, Schneider P, Wissemann V, Stuppner H.* Metabolic fingerprinting of *Leontopodium* species (Asteraceaea) by means of ^{1}H-NMR and HPLC-ESI-MS. Phytochemistry 2011; 72: 1379–1389

18 *Yilmaz A, Nyberg N, Mølgaard P, Asili J, Jaroszewski J.* ^{1}H NMR metabolic fingerprinting of saffron extracts. Metabolomics 2010; 6: 511–517

19 *Ali K, Maltese F, Zyprian E, Rex M, Choi YH, Verpoorte R.* NMR metabolic fingerprinting based identification of grapevine metabolites associated with downy mildew resistance. J Agric Food Chem 2009; 57: 9599–9606

20 *Kim HY, Choi YH, Erkelens C, Lefeber AW, Verpoorte R.* Metabolic fingerprinting of *Ephedra* species using ^{1}H-NMR spectroscopy and principal component analysis. Chem Pharm Bull (Tokyo) 2005; 53: 105–109

21 *Kim HK, Saifullah, Khan S, Wilson EG, Kricun SD, Meissner A, Goraler S, Deelder AM, Choi YH, Verpoorte R.* Metabolic classification of South American *Ilex* species by NMR-base metabolomics. Phytochemistry 2010; 71: 773–784

22 *Wold S.* Pattern-recognition by means of disjoint principal components models. Pattern Recogn 1976; 8: 127–139

23 *Wabuyele BW, Harrington PD.* Fuzzy optimal associative memory for background prediction of near-infrared spectra. Appl Spectrosc 1996; 50: 35–42

24 *Harrington PB, Kister J, Artaud J, Dupuy N.* Automated principal component-based orthogonal signal correction applied to fused near infrared-mid-infrared spectra of French olive oils. Anal Chem 2008; 81: 7160–7169

25 *Harrington PB.* Fuzzy multivariate rule-building expert systems – minimal neural networks. J Chemometr 1991; 5: 467–486

26 *CBOL Plant Working Group.* A DNA barcode for land plants. Proc Natl Acad Sci U S A 2009; 106: 12794–12797

27 *Gafner S, Blumenthal M, Reynaud DH, Foster S, Techen N.* ABC review and critique of the research article "DNA barcoding detects contamination and substitution in North American herbal products" by Newmaster et al. HerbalEGram 2013; 10: 11

28 *Reynaud D, Handy S.* Primers for short DNA sequences. AuthenTechnologies, LLC and US Food and Drug Administration. Personal communication; 2015

29 *Reynaud D.* DNA sequencing of SRM 3295, *Actaea racemosa*. AuthenTechnologies, LLC, Personal communication; 2014

30 *Strobel G, Daisy B.* Bioprospecting for microbial endophytes and their natural products. Microbiol Mol Biol Rev 2003; 67: 491–502

31 *Genazzani AD, Rodbard D.* Use of the receiver operating characteristic curve to evaluate sensitivity, specificity, and accuracy of methods for detection of peaks in hormone time-series. Acta Endocrinol (Copenh) 1991; 124: 295–306

32 *Hanley JA, Mcneil BJ.* The meaning and use of the area under a receiver operating characteristic (Roc) curve. Radiology 1982; 143: 29–36

33 *Harrington PDB.* Statistical validation of classification and calibration models using bootstrapped Latin partitions. Trends Anal Chem 2006; 25: 1112–1124

34 *Sun XB, Zimmermann CM, Jackson GP, Bunker CE, Harrington PB.* Classification of jet fuels by fuzzy rule-building expert systems applied to three-way data by fast gas chromatography-fast scanning quadrupole ion trap mass spectrometry. Talanta 2011; 83: 1260–1268

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.