Plant Seed Species Identification from Chemical Fingerprints: A High-Throughput Application of Direct Analysis in Real Time Mass Spectrometry

Ashton D. Lesiak,[†] Robert B. Cody,[‡] A. John Dane,[‡] and Rabi A. Musah*,[†]

[†]Department of Chemistry, University at Albany, State University of New York, 1400 Washington Avenue, Albany, New York 12222, United States

[‡]JEOL USA Inc., 11 Dearborn Road, Peabody, Massachusetts 01960, United States

S *Supporting Information*

ABSTRACT: Plant species identification based on the morphological features of plant parts is a well-established science in botany. However, species identification from seeds has largely been unexplored, despite the fact that the seeds contain all of the genetic information that distinguishes one plant from another. Using seeds of genus *Datura* plants, we show here that the mass spectrum-derived chemical fingerprints for seeds of the same species are similar. On the other hand, seeds from different species within the same genus display distinct chemical signatures, even though they may contain similar characteristic biomarkers. The intraspecies chemical signature similarities on the one hand, and interspecies fingerprint differences on the other, can be processed by multivariate statistical analysis methods to enable rapid species-level identification and differentiation. The chemical fingerprints can be acquired rapidly and in a high-throughput manner by direct analysis in real time mass spectrometry (DART-MS) analysis of the seeds in their native form, without use of a solvent extract. Importantly, knowledge of the identity of the detected molecules is not required for species level identification. However, confirmation of the presence within the seeds of various characteristic tropane and other alkaloids, including atropine, scopolamine, scopoline, tropine, tropinone, and tyramine, was accomplished by comparison of the in-source collision-induced dissociation (CID) fragmentation patterns of authentic standards, to the fragmentation patterns observed in the seeds when analyzed under similar in-source CID conditions. The advantages, applications, and implications of the chemometric processing of DART-MS derived seed chemical signatures for species level identification and differentiation are discussed.
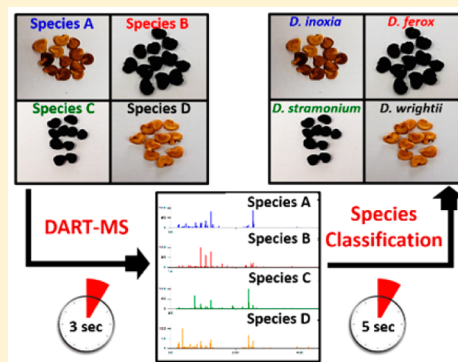
Plant seeds are not only a receptacle for a unique genome that distinguishes the embryonic plant from every other plant species but also a reservoir for a tailored mix of the nutrients required to support plant germination and early growth for the particular species. The outward manifestation of the genomic and nutrient requirement distinctions between seeds of individual species takes the form of differences in chemical profiles. As such, seeds would be expected to exhibit a metabolome that is similar for plants within the same species, but different between species. Furthermore, unlike the metabolic profile of the organs in a germinated plant that reflect significant real-time fluctuations in gene expression, and which in turn introduce small-molecule fingerprint variations, such gene expression variations may occur to a much lower extent in seeds that are not undergoing germination. This would be expected to result in much less variability in the chemical signatures of seeds of the same species when compared to what would be observed in examinations of plant aerial parts or roots. Therefore, seed metabolome fingerprinting and profiling potentially present useful, largely unexploited opportunities in several areas, including seed species identification, assessment of seed development stages,

and determination of seed exposure to pests, disease, or chemical agents. However, relatively few seed metabolome studies have been reported, in contrast to proteomics investigations. Those that have appeared focus by and large on agriculturally important food crops and include: (1) investigations of metabolite variations among cultivars within a single species;[1−3] (2) the monitoring of plant biosynthesis reactions in developing and germinating seeds;[4−6] (3) identification and quantification of medicinal natural products such as glucosinolates;[7] and (4) assessment of small-molecule biomarker variations in seeds in response to genetic manipulation.[8−10] Seed metabolome profiling has also been exploited in forensics applications, such as in the identification of cannabis seeds.[11]

The steps associated with most seed metabolome determination studies begin with solvent extraction. Depending on the profile of compounds known or expected to be present, the

extract may be subjected to further treatment such as the addition of derivatizing agents to enhance the volatility of analytes and pH adjustments followed by filtration to remove constituents whose presence is incompatible with subsequent analysis steps. The resulting solution is then subjected to analysis by a spectroscopic method which may or may not be hyphenated (e.g., NMR, Raman, IR, LC-FT-IR, LC-NMR) and/or a hyphenated mass spectrometric method [e.g., GC/MS, LC-MS, and capillary electrophoresis mass spectrometry (CE-MS) among others]. GC/MS and the hyphenated methods offer the advantage that complex mixtures can be separated into constituent components that can then be analyzed structurally to obtain more definitive information about specific compounds. Whatever the method used, the observed metabolome fingerprint can be processed by various multivariate statistical methods to gain insights into a variety of processes and their interrelatedness. The experimental protocols associated with acquisition of metabolite fingerprints by these conventional methods can be tedious and time-consuming, particularly with regard to sample preparation. Furthermore, extraction solvent choice is extremely important because the differential uptake of compounds by the solvent influences the profile of molecules detected. An additional challenge is that chemometric processing of the metabolome profile/fingerprint is hampered by the large number of replicates required to assess statistical significance, coupled with the considerable time that is necessary to acquire a sufficient number of them.

One way to circumvent the aforementioned bottlenecks associated with metabolic profiling by conventional methods is to use the solvent-free approach afforded by direct analysis in real time mass spectrometry (DART-MS). By this ambient ionization MS method, solids, liquids, or gases can be analyzed directly by exposing them to the open air space between the mass spectrometer inlet and the ion source.[12] A mass spectrum is acquired instantly, which provides high-throughput opportunities. The technique has been demonstrated to detect a range of both polar and nonpolar analytes in complex matrices including plant material of various forms.[12−15] Thus, analysis by this method has been demonstrated to provide chemical fingerprints that can be subjected to analysis by multivariate statistical methods. For example, statistical analysis of DART-MS derived fingerprints has been used to distinguish between wood types,[13] identify the country of origin of medicinal plants,[16] distinguish between organic and conventional farming methods in the cultivation of economically important crops,[17] and differentiate between propolis of various origins.[18]

In our ongoing work on the development of methods that can be used to rapidly identify plant species, we sought to determine whether the DART-MS-derived chemical profile of seeds could be used to differentiate between plant species within the same genus. For the study, we selected *Datura* seeds as a convenient example of a genus that contains a finite number of species most of which are available from commercial sources, namely, *D. ceratocaula*, *D. discolor*, *D. ferox*, *D. inoxia*, *D. leichhardtii*, *D. metel*, *D. quercifolia*, *D. stramonium*, and *D. wrightii*. *Datura* spp. are important in horticulture as ornamentals. Plants of this genus also contain toxic hallucinogens including scopolamine and atropine. Historically, the seeds and aerial plant parts have been used in ceremonial rituals to induce mind-altering effects. It is often difficult to identify *Datura* species because not only are the plants similar in appearance but also their morphological features often

resemble those of plants in the *Brugmansia* genus.[19] In addition, the seeds of several of the species are very similar, making it difficult to tell them apart (Figure 1). *D. ceratocaula*, *D. discolor*,
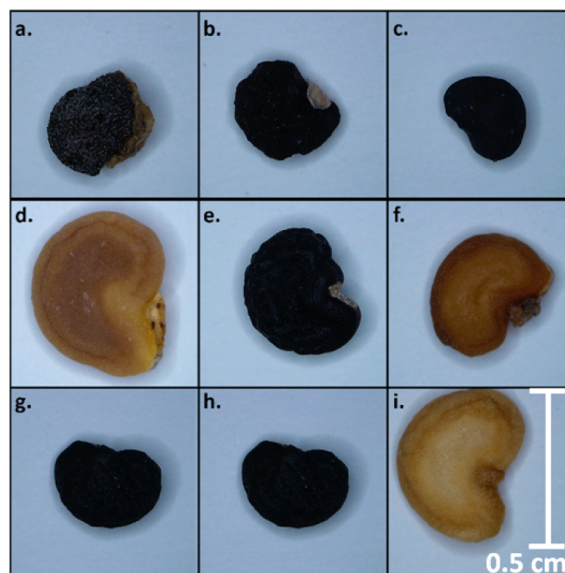
Figure 1. Light microscopy images of *Datura* spp. seeds. Panel a: *D. ceratocaula*; Panel b: *D. discolor*; Panel c: *D. ferox*; Panel d: *D. inoxia*; Panel e: *D. leichhardtii*; Panel f: *D. metel*; Panel g: *D. quercifolia*; Panel h: *D. stramonium*; Panel i: *D. wrightii*.

D. ferox, *D. leichhardtii*, *D. quercifolia*, and *D. stramonium* seeds are black in color, while those of *D. inoxia*, *D. metel*, and *D. wrightii* are light brown. Significantly, it is sometimes not until *Datura* plants reach full maturity that it is possible to identify the species based on the appearance and morphology of the seedpods.[19] This has led to misidentification of some genus *Brugmansia* plants as belonging to the *Datura* genus and vice versa. Therefore, the ability to differentiate and classify *Datura* species would serve an agronomic purpose. Moreover, it would be useful to identify *Datura* seed material in a forensics context, as abuse of the plant is on the rise and there are few protocols available for its definitive identification.[20,21]

In previously reported analyses of *Datura* alkaloids, seeds were ground and the resulting pulp extracted for a period of up to 18 h[22,23] (or extracted for shorter time periods but with multiple replicates).[24] Depending upon the study, the extracts were then subjected to a solvent removal step (e.g., lyophilization or evaporation),[22−24] a removal of plant oils step, reconstitution of the residue in the analysis solvent, and in some cases, a derivatization step to increase the volatility of analytes of interest.[22,23,25] Samples thus prepared were then subjected to spectroscopic analysis by NMR, Raman, or FT-IR or hyphenated mass spectrometric methods such as GC/MS, LC-MS, or CE-MS. Although these methods do allow for separation of the mixtures into individual components, the developed protocols take from 20 to 60 min per sample[26−31] in addition to the time-consuming sample preparation steps. Moreover, the development of such protocols for each species within a genus of seeds is impractical. Genomic profiling through DNA analysis has been shown to enable identification of some seed species.[32−34] However, since the genomes of the majority of plants have not been mapped, this method can only be applied to the small number of mainly cash crops for which
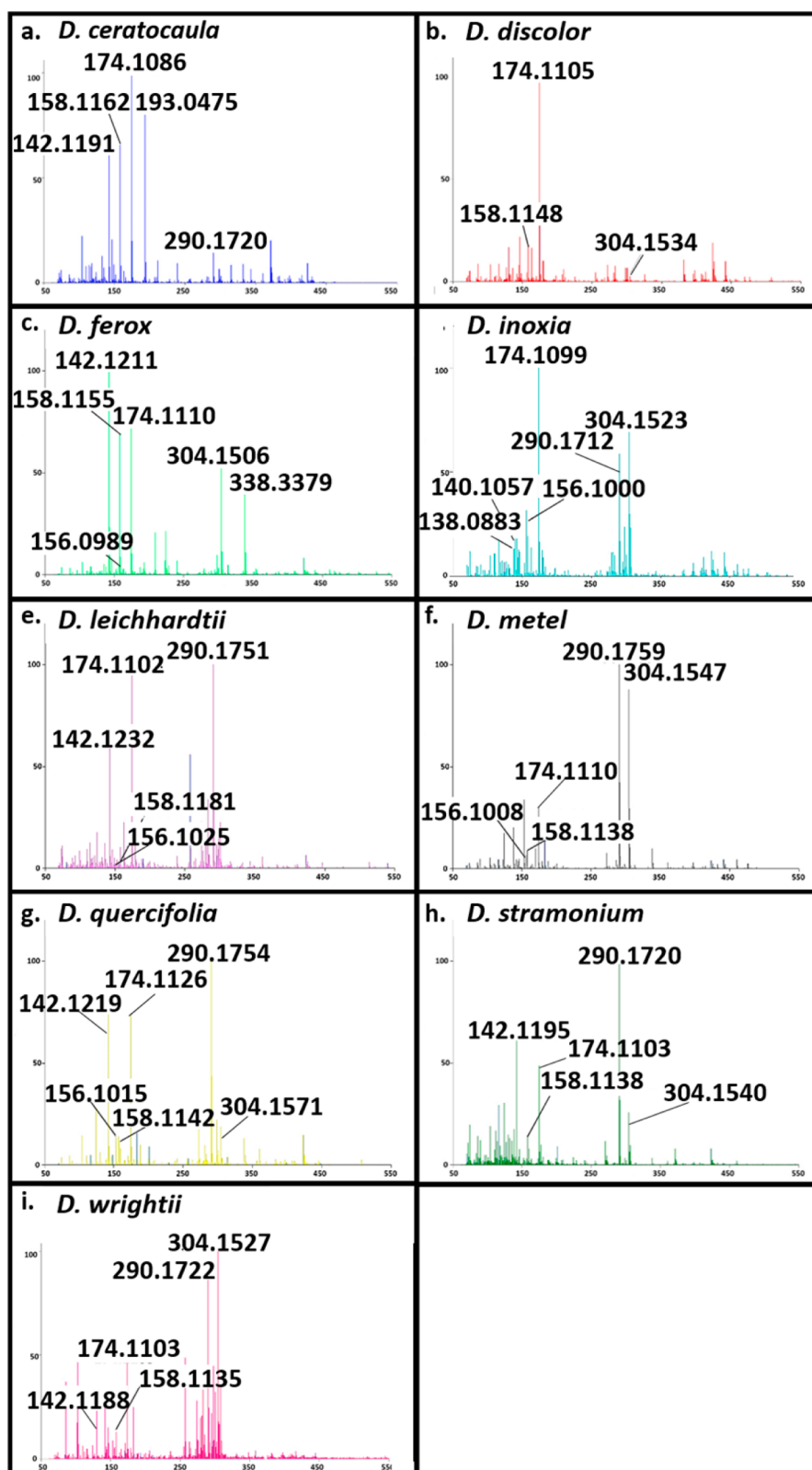
Figure 2. Positive-ion mode DART-HR-TOF-MS spectra of the seeds of nine species of the *Datura* genus. Panel a: *D. ceratocaula*; Panel b: *D. discolor*; Panel c: *D. ferox*; Panel d: *D. inoxia*; Panel e: *D. leichhardtii*; Panel f: *D. metel*; Panel g: *D. quercifolia*; Panel h: *D. stramonium*; Panel i: *D. wrightii*. The soft ionization spectra were collected at an orifice voltage of 20 V. They show the detected molecules in the seed in their protonated form. The spectra representing each species are visually distinct, indicating that the differences between the species could be exploited for class differentiation. Mass spectral data associated with the peaks this figure are listed in Table S-1.

gene sequence information is available (e.g., cereal grains and some vegetables).[32−42]

In this report, we describe the results of our investigation into whether DART-MS analysis could provide diagnostic

C

fingerprint profiles of *Datura* species seeds and whether chemometric processing of the observed profiles could enable species level identification and differentiation. We observe that the seeds can be analyzed by DART-MS directly in a high-throughput manner without the use of a solvent extract. Each species exhibits a distinct chemical signature, and the processing of this data by multivariate statistical methods enables species level differentiation. The results show that chemical fingerprinting of seeds is a viable method for species identification and differentiation. The protocol reported here circumvents bottlenecks commonly encountered in metabolome and fingerprint profiling investigations conducted by conventional methods.

■ EXPERIMENTAL SECTION

Seed Materials. When possible, seeds of a given species were acquired from multiple vendors. Thus, nine species of *Datura* seeds were purchased from a variety of vendors. *D. ceratocaula, D. discolor, D. inoxia, D. leichhardtii, D. stramonium,* and *D. wrightii* seeds were purchased from J.L. Hudson, Seedsman (La Honda, CA, USA). *D. ceratocaula, D. ferox, D. inoxia, D. metel, D. stramonium,* and *D. wrightii* seeds were purchased from Georgia Vines (Claxton, GA, USA). *D. ceratocaula, D. discolor,* and *D. ferox* seeds were purchased from World Seed Supply (Mastic Beach, NY, USA). *D. inoxia* and *D. stramonium* seeds were purchased from Horizon Herbs (Williams, OR, USA). *D. quercifolia* seeds were purchased from Hirts Gardens (Medina, OH, USA).

Standards. Atropine, scopolamine, tropinone, and tyramine standards were purchased from Sigma-Aldrich (St. Louis MO, USA). Scopoline was purchased from AK Scientific (Union City, CA, USA), and tropinone was purchased from Alfa Aesar (Ward Hill, MA, USA).

Microscopic Imaging. *Datura* spp. seeds were viewed using a Nikon stereozoom SMZ800 microscope that was equipped with a Nikon DS Fi2 microscope camera.

Vacuum Tweezer Apparatus. A vacuum tweezer apparatus was used to hold each analyzed seed half in the open air gap between the mass spectrometer inlet and the DART ion source (Figure S-1). The tweezer was formed as follows: a Wiretrol II disposable pipet capillary tube open at both ends (Drummond Scientific, Broomall, PA) was inserted into the smallest opening of a 200 μL Fisherbrand RediTip pipet tip (Fisher Scientific, Ontario, Canada), and a flame was applied briefly to the tapered end of the plastic tip in order for the melted plastic to form an airtight seal around the capillary. The opposite open end of the pipet tip was then attached to one end of a high temperature silicone rubber tube (1/4 in. inner diameter, 1/2 in. outer diameter) (McMaster-Carr, Robbinsville, NJ, USA), and the opposite end of the tube was connected to an MZ 2 diaphragm vacuum pump (Vacuubrand, Wertheim, Germany). Turning on the vacuum pump created suction in the capillary tube, which was then applied to the seed half to hold it steadily at the MS inlet during analysis. More than 100 seeds could be sampled per hour using the vacuum tweezer apparatus.

Mass Spectral Data Collection and Analysis. Mass spectra were acquired using a DART-SVP ion source (IonSense, Saugus, MA, USA) coupled to a JEOL AccuTOF time-of-flight mass spectrometer (JEOL USA, Peabody, MA, USA) in positive ion mode. The DART ion source parameters were: grid voltage, 250 V; gas heater temperature, 350 °C. Unless otherwise indicated, the mass spectrometer settings were: ring lens voltage, 5 V; orifice 1 voltage, 20 V; orifice 2 voltage, 5 V; peak voltage 600 V. Spectra were obtained over the mass range of m/z 60–800 at 1 spectrum per s. The helium flow rate for the DART source was 2.0 L s^{-1}. The resolving power of the mass spectrometer was 6000 fwhm.

Individual seeds of *Datura* species were sliced in half and sampled directly by gripping each sample with the vacuum tweezer apparatus (Figure S-1) and suspending the seed between the ion source and mass spectrometer inlet. This yielded the results presented in Figure 2. Each of the spectra shown is an average of three seed spectra (each one from a different vendor), except in the cases of *D. leichhardtii, D. metel,* and *D. quercifolia,* where seeds were available from only one vendor. For these species, three spectra from the single vendor were averaged. Seed extracts were prepared by slicing a seed in half and suspending both halves in 50 μL of solvent and sonicating for 30 min. Solvents used were hexane, ethyl acetate, dichloromethane, ethanol, and water. A 1:1:1 mixture of ethyl acetate, ethanol, and water was also used. Extracts were tested directly by dipping the closed end of a melting point capillary tube into the extract and suspending the coated surface of the tube between the DART ion source and the mass spectrometer inlet.

In-source collision-induced dissociation (CID) was performed on seed material and standards by using "function switching", which permits the simultaneous acquisition of both low orifice voltage and high orifice voltage mass spectra. To accomplish this, the orifice 1 voltage was varied from 20 to 30, 60, and 90 V, with the extent of fragmentation increasing with increasing voltage.[43,44] All other DART-SVP and AccuTOF parameters remained as described above. Chemical standards were tested directly by dipping the closed end of a melting point capillary tube into the pure standard and suspending the coated surface of the tube between the DART ion source and the mass spectrometer inlet. Seeds were analyzed in the manner described above.

Data Processing. Calibration, spectral averaging, background subtraction, and peak centroiding were performed using TSSPro3 (Shrader Analytical Laboratories, Detroit, MI, USA) data processing software. Mass Mountaineer software (Massspec-software.com, Toronto, Ontario, Canada) was used for mass spectrum analysis, spectral elemental composition determination, isotope analysis, and spectra addition, as well as for linear discriminant analysis (LDA)-facilitated classification and discrimination. A threshold of 10% was used for reporting mass spectral data, except in the cases of the eight diagnostic alkaloids for the *Datura* spp. (tyramine, tropinone, tropine, scopoline, dihydroxytropane, trihydroxytropane, atropine/hyoscyamine, and scopolamine), which were reported using a 2% threshold (Tables S-1 and S-12). The 10% threshold was used in order to reduce to an appropriate size (for reporting purposes) the large amount of mass spectral peak data that were acquired (although all the spectra are shown). The normalized relative abundance values of the eight diagnostic alkaloids were calculated and averaged across each species. Relative standard deviation values were calculated from the averaged abundances and the standard deviation values of these alkaloids for each species. Mass calibration was performed using polyethylene glycol (PEG 600) or Jeffamine M600 (Huntsman, The Woodlands, TX). For readability, nominal masses are listed throughout the text. Accurate masses are reported in the mass spectra and in the Supporting Information.

■ RESULTS

DART-MS Analysis of Seeds. To determine the chemical fingerprints of the seeds of various *Datura* spp., seeds representing 9 species were analyzed by DART high resolution time-of-flight mass spectrometry (HR-TOF-MS). These were *D. ceratocaula*, *D. discolor*, *D. ferox*, *D. inoxia*, *D. leichhardtii*, *D. metel*, *D. quercifolia*, *D. stramonium*, and *D. wrightii*. In every case, 50 seeds were analyzed in positive-ion mode. To eliminate potential bias associated with analysis of seeds that were acquired from a single source, seeds of the same species were acquired from as many vendors as were available. Each seed was cut in half, and the halves were sampled by suspending them using a vacuum tweezer apparatus (described in the Experimental Section) in the open air space between the DART ion source and the mass spectrometer inlet, with the sliced end of the seed facing the ion source. Representative spectra obtained in each case are shown in Figure 2, and the corresponding mass spectral data including measured and calculated masses and relative peak abundances are presented in Table S-1. Each of the images in Panels a through i represents the average of 3 spectra of a given species. The spectra contained from 52 to 117 peaks (using a 2% threshold relative to the base peak) with *D. quercifolia* (Panel g) exhibiting the greatest number of peaks and *D. ferox* the least (Panel c). The peaks observed in each spectrum represented the protonated forms of the detected molecules.

Visual inspection showed the fingerprint of each species to be distinct. Nevertheless, all of the spectra exhibited peaks at m/z values having molecular formulas consistent with those of several compounds that have been isolated from *Datura* species (Table S-2).[20,22,45−47] Prominent among them were psychotropic amine biomarkers and other related compounds including tyramine (m/z 138), tropinone (m/z 140), tropine (m/z 142), scopoline (m/z 156), dihydroxytropane (m/z 158), trihydroxytropane (m/z 174), scopoletine (m/z 193), and hydroxydiethyloxytropane (m/z 338). Also observed in every case, except for *D. discolor* and *D. ferox*, was a peak at m/z 290 corresponding to the formula $C_{17}N_{23}NO_3$. This formula is consistent with that of the psychotropic drug atropine, which has been identified in *Datura* spp. However, hyoscyamine, which is the enantiomer of atropine, has also been isolated from *Datura* genus plants, and thus, it is not certain which of the two molecules is represented. A peak at m/z 304, which corresponds to the molecular formula $C_{17}H_{21}NO_4$, was observed in all species except *D. ceratocaula* and *D. leichhardtii*. The compound was tentatively identified as scopolamine/hyoscine. This also represents an enantiomeric pair, and both molecules have been isolated from *Datura* spp.[22,46,48] The relative abundances of the m/z values representing the various tropane alkaloids observed for each species are presented in Table S-2.

The data show that the relative proportions of these biomarkers differed from species to species. Some species exhibited comparable levels of both atropine/hyoscyamine (m/z 290) and scopolamine/hyoscine (m/z 304) (i.e., *D. ferox*, *D. inoxia*, *D. metel*, and *D. wrightii*; Panels c, d, f, and i, respectively). Others had much higher levels of atropine/hyoscyamine than scopolamine/hyoscine (i.e., *D. leichhardtii*, *D. quercifolia*, and *D. stramonium*) (Panels e, g, and h). Scopolamine/hyoscine was most prominent in *D. ferox*, *D. inoxia*, and *D. wrightii* (Panels c, d, and i). In several species, various tropane alkaloids were prominent enough to be the

base peak (atropine/hyoscyamine in *D. leichhardtii*, *D. metel*, *D. quercifolia*, and *D. stramonium* (Panels e, f, g, and h); scopolamine/hyoscine in *D. wrightii* (Panel i; and trihydroxytropane in *D. ceratocaula*, *D. discolor*, and *D. inoxia* (Panels a, b, and d)). *D. ceratocaula* and *D. discolor* exhibited particularly low levels of atropine/hyoscyamine and scopolamine/hyoscine compared to the other species analyzed (Panels a and b).

In-Source Collision-Induced Dissociation (CID) Experiments Confirmed the Identity of Various Tropane Alkaloids in *Datura* spp. DART-MS analysis is a soft ionization technique that results in the detection primarily of protonated molecules when conducted in positive ion mode at a low orifice voltage (e.g., 20 V). However, DART-MS can be conducted under in-source collision-induced dissociation (CID) conditions to yield fragmentation patterns that serve as a signature for the analyzed molecule. We exploited this technique to confirm the presence in *Datura* species of alkaloids for which we had authentic standards. In-source CID was accomplished by increasing the orifice voltage during the analysis. In order to determine the optimal voltage at which to conduct the in-source CID experiments, authentic standards of atropine, scopolamine, scopoline, tropine, tropinone, and tyramine were analyzed by DART at orifice voltages of 20, 30, 60, and 90 V. The 90 V condition was consistently observed to yield the best results (i.e., maximum fragmentation with retention of the protonated parent molecule peak). The DART mass spectra of the seeds of each species were then determined at 90 V, and the spectrum in each case was compared to the spectrum determined at 90 V for each of the available standards. Representative results are shown for *D. metel* in Figure S-2, and a list of the mass spectral data associated with each of the alkaloids whose presence in *Datura* spp. seeds was confirmed appears in Table S-2. Each of the Panels a through f shows two in-source CID spectra rendered as head-to-tail plots. The top spectrum in each plot is that of the seed at 90 V, and the bottom spectrum is that of the indicated standard at 90 V. Thus, the results shown in Panel a illustrate that all of the atropine fragments observed from in-source CID of the atropine standard appear in the spectrum of *D. metel*, thus confirming that the peak at m/z 290 represents atropine. A similar result was obtained for scopolamine, as shown in Figure S-2, Panel b. The 90 V spectrum of *D. metel* shows not only the parent scopolamine but also the scopolamine fragments, confirming that the peak at m/z 304 is scopolamine. In Figure S-2, Panel c, the head-to-tail plot shows the 90 V spectrum of the *D. metel* seed on top and the 90 V spectrum of scopoline on the bottom. The fragments at m/z 138, 110, 94, and 67, as well as the protonated parent at m/z 156, match in both the standard and the seed, confirming the presence of scopoline in the seed material. Similarly, in Panel d of Figure S-2, the in-source CID spectrum of the seed shows peaks that match those of the tropine standard (at m/z 158, 142, 124, 93, and 67). This confirmed the presence of tropine in *D. metel* seeds. Panel e shows the head-to-tail comparison spectra of the *D. metel* seed and tropinone standard. Peaks at m/z 156, 140, and 82 in both the seed material and standard confirm that tropinone is present in the seed. Tyramine was confirmed to be present in the *D. metel* seed through in-source CID analysis as well. In Panel f, the head-to-tail plot shows the 90 V spectrum of *Datura* seed on top and the 90 V spectrum of tyramine on the bottom. The peaks at m/z 138, 121, 103, 91, 77, and 63 that are present in the standard are also present in the seed material, confirming the presence of tyramine.

DART-MS Analysis of Seed Extracts. Conventional methods of metabolome or chemical fingerprint determination often rely on the analysis of extracts of the plant material rather than direct analysis of the material itself. For this reason, it was of interest to compare the results obtained by direct analysis of seeds as described above, to those obtained by DART-MS analysis of seed extracts. Extracts of each *Datura* seed species were prepared using five solvents, namely, hexane, ethyl acetate, dichloromethane, ethanol, and water. Extracts were prepared by halving a seed, suspending both halves in 50 μL of solvent, and sonicating for 30 min. The resulting solution was then immediately analyzed by DART-MS. The spectra obtained are shown in Figures S-3–S-11. Examination of the data revealed a number of trends. As expected, a select group of components was detected in each solvent and there was very little similarity between the extract spectra. A representative example of the typical differences seen between the DART-MS-derived fingerprints of the extracts for a given species is shown for *D. metel* (Figure S-3 and Table S-3).

Hexane proved to be the least effective extraction solvent based upon the greatly reduced number of detected molecules compared with the number detected in other solvents (Figures S-3–S-11). Moreover, the diagnostic tropane alkaloids were never detected in the hexane or dichloromethane extracts in amounts greater than 2% (Figures S-3–S-11, Panels a and c). Ethyl acetate, ethanol, and water were all effective in extracting the tropane alkaloids (Figures S-3–S-11, Panels b, d, and e). Nevertheless, the fingerprints of these three extracts were visually very different. The distinctions were a consequence not only of differences in the relative abundances of peaks common to all three spectra, but also because the profiles of extracted molecules differed between solvents. For example, the base peak in the ethyl acetate spectrum was at m/z 195. However, this peak, which was also the base peak in the spectra of the hexane and dichloromethane extracts, did not appear at all in the ethanol and water extracts. Comparison of each of the extract spectra in Figure S-3 to that of the spectrum obtained from direct analysis of the seed (Figure 2, Panel f) showed the latter to provide the most comprehensive fingerprint profile. Similar trends were observed in comparisons between extracts for the other analyzed species (Figures S-3–S-11 and Tables S-3–S-11). Thus, for all analyzed species, the base peak in the hexane and dichloromethane extracts was always at m/z 195. This peak was never present in the spectra of the ethanol and water extracts. The identity of the molecule represented by this peak is unknown.

Comparisons between the DART-MS Spectra of Seeds and Those of Extracts Revealed That Direct Seed Analysis Was Superior for the Detection of Alkaloid Biomarkers. Comparison of the spectra obtained from direct seed analysis to those of the extracts revealed trends that were similar across the range of analyzed species. The case of *D. ceratocaula* is representative (Figure 2, Panel a). The most prominent peaks in the seed spectrum were those of the alkaloid biomarkers tropine (m/z 142, 68.5% relative abundance), dihydroxytropane (m/z 158, 65.6% relative abundance), trihydroxytropane (m/z 174, 100% relative abundance), scopoletin (m/z 193, 81% relative abundance), and atropine (m/z 290, 14.4% relative abundance). With the exception of trace to modest relative amounts of atropine detected in the ethyl acetate, dichloromethane, ethanol, and water extracts (0.4%, 0.4%, 0.3%, and 26.1% relative abundances, respectively) and a relatively low amount of

scopolamine detected in the water extract (1.7% relative abundance), none of the other diagnostic biomarkers were observed in the solvent extracts (Figure S-4 and Table S-4)

In *D. inoxia* (Figure 2, Panel a), seven diagnostic alkaloids were detected by direct seed analysis, namely, tyramine (m/z 138, 13.2% relative abundance), tropinone (m/z 140, 28.2% relative abundance), scopoline (m/z 156, 31.8% relative abundance), trihydroxytropane (m/z 174, 100% relative abundance), atropine (m/z 290, 61.9% relative abundance), and scopolamine (m/z 304, 70% relative abundance). With the exception of the observation of scopolamine in the ethyl acetate, dichloromethane, ethanol, and water extracts (at 17.5%, 1.9%, 38.4%, and 100% relative abundances, respectively) and detection of tyramine, trihydroxytropane, and scopolamine in the water extract (at 52.5%, 80.6%, and 100% relative abundances, respectively), the alkaloids detected in the seed were absent from the extracts (Figure S-7 and Table S-7). Comparison of the seed DART-MS spectra (Figure 2) to those of the corresponding seed extracts (Figures S-3–S-11) revealed analogous trends in all cases. Direct seed analysis was consistently observed to reveal the presence of the most comprehensive set of alkaloid biomarkers. Although the aqueous extract of every species of seed always contained alkaloid secondary metabolites, the entire complement of the amines detected by direct seed analysis was never observed in the water extracts. However, the solvents did exhibit differential uptake of various alkaloids. For example, the ethyl acetate extract of *D. metel* contained atropine and scopolamine but no tyramine (Figures S-3). On the other hand, the ethanol extract showed the presence of atropine, scopolamine, and tyramine, and the water extract showed four alkaloids, namely, tyramine, trihydroxytropane, atropine, and scopolamine. In order to determine whether a solvent extract could yield a fingerprint that was more aligned with that observed by direct analysis of the seeds, a three-component solvent mixture comprised of a 1:1:1 ratio by volume of ethyl acetate, ethanol, and water was used to extract the seeds of each species. The three solvents in the mix were chosen because all three demonstrated the ability to take up various alkaloids, while hexane and dichloromethane did not. The DART mass spectrum obtained for each species is shown in Figure S-12 with the mass spectral peak information presented in Table S-12. A representative example of the result for *D. metel* is shown in the head-to-tail plot in Figure 3. The top panel of the plot shows the DART mass spectrum obtained from direct analysis of the seed, and the bottom panel shows the results for the 3-solvent extract. Remarkably, the spectra are very similar both in terms of the m/z values observed and their relative intensities. This demonstrated that it was possible for an appropriately chosen solvent system to extract compounds that are similar to those that are observed by direct seed analysis. However, when taking into account time per analysis, as well as the process of determining the most appropriate extraction solvent system, direct seed analysis is superior, as it provides similar results without the added time needed for extraction and method development.

Statistical Analysis of Seed DART-MS Derived Chemical Fingerprints Enabled Species-Level Identification. To determine whether statistical analysis could be used to classify and correctly identify seed species based on their DART-MS-derived mass spectral fingerprints, a training set of 121 mass spectra representing all 9 *Datura* species was employed. From the spectra, 35 feature masses (listed in Table S-13) were selected for linear discriminant analysis
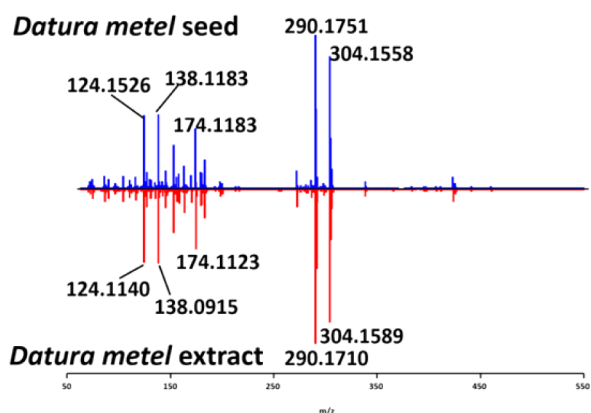
Figure 3. Positive-ion mode DART-HR-TOF-MS spectra of a *D. metel* seed compared to that of a *D. metel* extract made using a solvent comprised of 1:1:1 mixture v/v of ethyl acetate, ethanol, and water, rendered as a head-to-tail plot. The seed spectrum is shown at the top and the extract spectrum on the bottom. The two spectra are very similar, indicating that the direct analysis of seed material yields results comparable to those observed by analysis of the indicated extract.

(LDA). The LDA score plot results are shown in Figure 4. The panels in Figure 4 show different orientations of the plot in order to illustrate the clustering of species and the separation between clusters. Panel a shows that the *D. ceratocaula*, *D. discolor*, *D. ferox*, and *D. leichhardtii* clusters are well resolved from one another with no overlap between these species. The *D. ceratocaula* and *D. discolor* clusters are lower on the *y*-axis than the other clusters. Panel b shows another orientation of the same LDA plot and illustrates that the *D. ferox*, *D. ceratocaula*, and *D. discolor* clusters are well resolved from the other species. Furthermore, this view illustrates that the *D. stramonium* and *D. quercifolia* clusters do not to overlap with the *D. wrightii* or *D. ferox* clusters, as was implied in Panel a. From the vantage point in Panel c, the resolution of the *D. inoxia*, *D. metel*, and *D. wrightii* clusters is apparent. Panel d shows a top down image of the LDA plot with the *D. ceratocaula* cluster removed to provide clarity. From this view, it can be seen that the data points in the *D. stramonium* and *D. quercifolia* clusters are resolved. This panel also shows that the *D. inoxia*, *D. metel*, and *D. wrightii* clusters are separated and that the *D. leichhardtii* cluster is resolved from those of the other species. Thus, the LDA scatter plot demonstrated that statistical processing of the DART-MS data permitted species
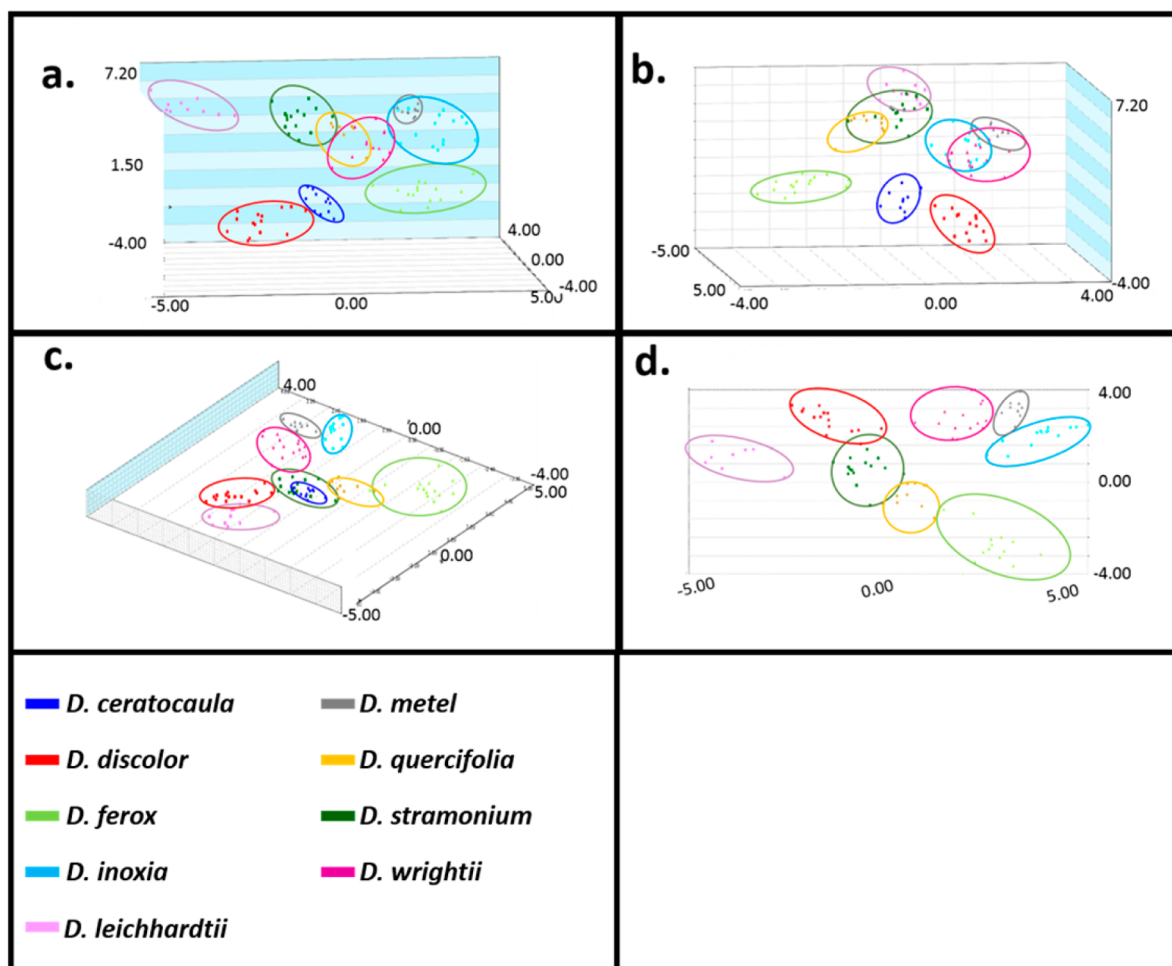
Figure 4. Linear discriminant analysis (LDA) plot of mass spectral data derived from DART-HR-TOF-MS analysis of *Datura* spp. seeds using the 35 feature masses listed in Table S-13. Panels a through d show different orientations of the same plot in order to better illustrate the clustering observed for spectra representing the same species, and the resolution of clusters representing different species from one another. Eight principal components covered 98.02% of the variance, and the leave-one-out-cross-validation (LOOCV) was 89.26%. The various colors correspond to the indicated species.

G

level differentiation. Moreover, eight principal components (PCs) accounted for 98.02% of the variance, and the leave-one-out cross validation (LOOCV) was 89.26%.

■ DISCUSSION

On the basis of the premise that the chemical profiles of seeds of the same species are likely to be consistently similar in a way that enables them to be distinguished from those of a different species, we launched an investigation into whether the mass spectrum-derived chemical fingerprint of seeds could be used for plant species identification. The ability to identify species from plant seeds could potentially have far reaching implications in several fields including botany, agriculture, and forensics. It could permit differentiation of species of plants within the same genus, identification of genetically modified crops from seeds, and evaluation of species relationships and sources of food stocks.[49]

The availability of *Datura* spp. seeds presented an optimal opportunity for us to test our hypothesis, as the genus contains only a small number of recognized species (<10) and it is well-characterized by conventional analytical methods due to its relevance as an ornamental plant and use as a psychoactive drug.[19,21] The limited number of species enabled us to obtain a comprehensive mass spectral data set for the available recognized species so that similarities and differences between them could be readily detected. Historically, it is has proven difficult to distinguish between *Datura* species that have not attained maturity due to similarities in the appearance of both their seeds and aerial parts and because their morphological features can vary depending on where the plants are grown.[19] Furthermore, most of the plants in this genus have been observed to contain tropane alkaloids, and thus, detection of these substances does not in and of itself enable identification of the specific species.

In the work presented here, the seed chemical fingerprints of nine *Datura* species were obtained readily by DART-HR-TOF-MS without the need for extraction, derivatization, or other steps described as necessary in previous studies.[22,45−47] The analysis was straightforward. A seed was simply cut in half and positioned in the space between the ion source and mass spectrometer inlet with the sliced end of the seed facing the ion source. This produced an almost instantaneous mass spectrum which showed a profile of the protonated forms of molecules detected in the seed, and multiple seeds could be sampled rapidly (approximately two seeds per minute). The accurate mass data enabled determination of molecular formulas that were consistent with those of biomarkers and other compounds that have been previously reported in *Datura* species.[20,22,46,47]

Seeds from multiple vendors were used in order to demonstrate the ability to distinguish between species despite different harvesting conditions. The chemical fingerprints of the *Datura* species were reproducible, even in comparisons of the seeds from different vendors, which enabled discrimination between species. Relative standard deviation (RSD) values associated with eight diagnostic alkaloid peaks were used to assess spectral reproducibility. These RSDs (calculated from 121 mass spectra from nine *Datura* species) are shown in Table S-14. The median RSDs ranged from 12.86% to 47.47%. Masses with the largest RSD values were from both high intensity (*m/z* 290, 304) and low intensity (*m/z* 140, 156) peaks, and thus, some intraspecies variation in the spectra was observed. We ascribe these differences to (1) the use of species from different sources, which were cultivated under different

plant growth conditions, and (2) seed sampling variations that were inherent in the manual method used to suspend the seeds between the ion source and mass spectrometer inlet. However, the presence of these spectral variations did not preclude discrimination between species based on the mass spectral profiles.

A benefit of this direct seed analysis sampling method is that no extraction protocol is needed, and therefore, no solvent selection bias is introduced into the results. Moreover, the analysis time is dramatically reduced compared to that typically required using conventional methods of alkaloid determination in *Datura* species.[22,25,45−48] The mass spectra of the individual species showed that the fingerprints are visually distinguishable, based on the number of peaks, the relative abundances of peaks, and their *m/z* values. For example, the spectrum of *D. ferox* exhibited only 52 peaks while the spectrum of *D. quercifolia* contained 117 peaks (using a 2% threshold to identify diagnostic *Datura* spp. alkaloids). Similarly, the abundances of certain characteristic alkaloids also varied greatly between species. For example, the base peaks in *D. stramonium* and *D. inoxia* were atropine/hyoscyamine (*m/z* 290) and scopolamine/hyoscine (*m/z* 304), respectively, while *D. discolor* and *D. inoxia* exhibited little to no scopolamine/hyoscine or atropine/hyoscyamine, respectively. It was the variation in the spectra between different species that enabled species discrimination. Our observations may explain the preference for some species of *Datura* seeds over others in a drug abuse context. *D. stramonium* and *D. inoxia* are two of the most commonly abused seeds in the genus, most likely due to their relatively high belladonna alkaloid content.

The ability to perform in-source collision-induced dissociation (CID) on samples and standards enabled us to confirm the presence of several alkaloids, specifically atropine, scopolamine, tyramine, tropine, tropinone, and scopoline without the need for chromatographic separation. This approach is analogous in some ways to that of identifying an analyte by comparing its EI fragmentation pattern to that of an authentic standard. However, important distinctions between the in-source CID experiments performed here and compound identification through comparisons to EI fragmentation patterns exist. For example, in contrast to tandem mass spectrometry where a single precursor ion is isolated and fragmented using a single clean collision gas, all detected ions are being activated simultaneously with in-source CID, with helium serving as the collision gas. This latter scenario has the potential to produce overlapping fragment ions derived from different molecules, simply by virtue of the large number of ions simultaneously produced. Nevertheless, although the intensity of a given fragment peak in DART in-source CID can be distorted because of the presence of isomeric fragments, the identification and confirmation of the alkaloids remains unaffected. The ability to confirm the presence of multiple alkaloids in a single seed analysis without sample preparation steps has not been reported previously and demonstrates the power and utility of DART-MS for rapid screening and compound identification.

The work presented here allows for the simultaneous monitoring of a range of seed compounds without the need for imprint steps such as those that have been used in desorption electrospray ionization (DESI)-MS analyses,[50−52] extractions, or solvents. Moreover, we show that direct seed analysis (Figure 2) gave a profile of compounds that was superior to that obtained from analysis using any of the single

solvent systems we investigated. DART-MS analysis of an extract prepared using the combination solvent system (i.e., 1:1:1 v/v/v of water/ethanol/ethyl acetate) did yield a spectrum very similar to that observed by direct seed analysis (Figures 3 and S-12). This further demonstrated that the extraction step was not required since direct seed analysis yielded the same information.

The high-throughput capability of DART-MS analysis made practical the rapid acquisition of multiple spectra that could be processed by multivariate statistical analysis methods, with >100 samples being analyzed per hour. The use of seeds for chemical phenotyping with multivariate statistical analysis is ideal, as the seed contains the entire genetic makeup that distinguishes the plant. This approach is novel, as most definitive identifications of species are accomplished through DNA analysis, which can be expensive and is limited to the small number of plants with mapped genomes. Processing of the chemical signature data by LDA illustrated the power of this approach to provide species-level discrimination between seeds. The methodology presented here demonstrates that a viable alternative to species differentiation using gene sequence information is the application of statistical analysis processing of seed chemical fingerprints. The method can be applied to a very large number of species even if little information is available on the plant's chemical makeup or the identities of any of its biomarkers.

CONCLUSION

We have demonstrated using *Datura* spp. seeds that the small-molecule fingerprint of seeds can be obtained by DART-HRMS. While intraspecies chemical signatures are similar, interspecies fingerprints are distinct enough to be discriminated using multivariate statistical analysis tools. The results shown here demonstrate the application and utility of DART-MS fingerprinting in seed species determination.

ASSOCIATED CONTENT

⑤ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.5b01611.

12 additional figures and 14 tables referenced in the text (PDF)

AUTHOR INFORMATION

Corresponding Author
*E-mail: rmusah@albany.edu. Fax: 518-442-3462.

Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENTS

REFERENCES

(1) Lin, H.; Rao, J.; Shi, J.; Hu, C.; Cheng, F.; Wilson, Z. A.; Zhang, D.; Quan, S. *J. Integr. Plant Biol.* **2014**, *56*, 826−836.

(2) Tsopmo, A.; Muir, A. D. *J. Agric. Food Chem.* **2010**, *58*, 8715−8721.

(3) Hu, C.; Shi, J.; Quan, S.; Cui, B.; Kleessen, S.; Nikoloski, Z.; Tohge, T.; Alexander, D.; Guo, L.; Lin, H.; Wang, J.; Cui, X.; Rao, J.; Luo, Q.; Zhao, X.; Fernie, A. R.; Zhang, D. *Sci. Rep.* **2014**, *4*, 5067.

(4) Auger, B.; Marnet, N.; Gautier, V.; Maia-Grondard, A.; Leprince, F.; Renard, M.; Guyot, S.; Nesi, N.; Routaboul, J.-M. *J. Agric. Food Chem.* **2010**, *58*, 6246−6256.

(5) Shu, X.-L.; Frank, T.; Shu, Q.-Y.; Engel, K.-H. *J. Agric. Food Chem.* **2008**, *56*, 11612−11620.

(6) Na Jom, K.; Frank, T.; Engel, K.-H. *Metabolomics* **2011**, *7*, 102−117.

(7) Hong, E.; Kim, S.-J.; Kim, G.-H. *Food Chem.* **2011**, *128*, 1115−1120.

(8) Levandi, T.; Leon, C.; Kaljurand, M.; Garcia-Cañas, V.; Cifuentes, A. *Anal. Chem.* **2008**, *80*, 6329−6335.

(9) Frank, T.; Röhlig, R. M.; Davies, H. V.; Barros, E.; Engel, K.-H. *J. Agric. Food Chem.* **2012**, *60*, 3005−3012.

(10) Piccioni, F.; Capitani, D.; Zolla, L.; Mannina, L. *J. Agric. Food Chem.* **2009**, *57*, 6041−6049.

(11) Ovenden, S. P. B.; Gordon, B. R.; Bagas, C. K.; Muir, B.; Rochfort, S.; Bourne, D. J. *Aust. J. Chem.* **2010**, *63*, 8−21.

(12) Cody, R.; Laramee, J.; Durst, H. D. *Anal. Chem.* **2005**, *77*, 2297−2302.

(13) Espinoza, E. O.; Lancaster, C. A.; Kreitals, N. M.; Hata, M.; Cody, R. B.; Blanchette, R. A. *Rapid Commun. Mass Spectrom.* **2014**, *28*, 281−289.

(14) Cody, R. *Anal. Chem.* **2009**, *81*, 1101−1107.

(15) Nilles, J.; Connell, T.; Durst, H. *Anal. Chem.* **2009**, *81*, 6744−6749.

(16) Fukuda, E.; Uesawa, Y.; Baba, M.; Suzuki, R.; Fukuda, T.; Shirataki, Y.; Okada, Y. *Nat. Prod. Commun.* **2014**, *9*, 1591−1594.

(17) Novotna, H.; Kmiecik, O.; Galazka, M.; Krtkova, V.; Hurajova, A.; Schulzova, V.; Hallmann, E.; Rembialkowska, E.; Hajslova, J. *Food Addit. Contam., Part A* **2012**, *29*, 1335−1346.

(18) Morlock, G. E.; Ristivojevic, P.; Chernetsova, E. S. *J. Chromatogr. A* **2014**, *1328*, 104−112.

(19) Preissel, U.; Preissel, H.-G. *Brugmansia and Datura: Angel's Trumpets and Thorn Apples*; Firefly Books: Richmond Hill, ON, 2002.

(20) Koverman, G. *Microgram* **1993**, *26*, 122.

(21) US Drug Enforcement Administration. *Jimson Weed (Datura stramonium)*; 2013; http://www.deadiversion.usdoj.gov/drug_chem_info/jimson_w.pdf. Accessed 7-14-2015.

(22) Friedman, M.; Levin, C. E. *J. Agric. Food Chem.* **1989**, *37*, 998−1005.

(23) Paul, B.; Dreka, C.; Knight, E.; Smith, M. *Planta Med.* **1996**, *62*, 544−547.

(24) Wu, X.; Li, N.; Li, H.; Tang, H. *Analyst* **2014**, *139*, 1769−1778.

(25) Namera, A.; Yashiki, M.; Hirose, Y.; Yamaji, S.; Tani, T.; Kojima, T. *Forensic Sci. Int.* **2002**, *130*, 34−43.

(26) Gorzolka, K.; Lissel, M.; Kessler, N.; Loch-Ahring, S.; Niehaus, K. *J. Biotechnol.* **2012**, *159*, 177−187.

(27) Kim, J. K.; Park, S.-Y.; Na, J.-K.; Seong, E. S.; Yu, C. Y. *J. Agric. Food Chem.* **2012**, *60*, 2257−2263.

(28) Matthews, S. B.; Santra, M.; Mensack, M. M.; Wolfe, P.; Byrne, P. F.; Thompson, H. J. *PLoS One* **2012**, *7*, e44179.

(29) Frank, T.; Rohlig, R.; Davies, H.; Barros, E.; Engel, K.-H. *J. Agric. Food Chem.* **2012**, *60*, 3005−3012.

(30) Sessa, R. A.; Bennett, M. H.; Lewis, M. J.; Mansfield, J. W.; Beale, M. H. *J. Biol. Chem.* **2000**, *275*, 26877−26884.

(31) Busko, M.; Jelen, H.; Goral, T.; Chmielewski, J.; Stuper, K.; Szwajkowska-Michalek, L.; Tyrakowska, B.; Perkowski, J. *Food Addit. Contam., Part A* **2010**, *27*, 1574−1581.

(32) Weder, J. K. P. *Lebensm. Wiss. Technol.* **2002**, *35*, 504−511.

(33) Devi, R.; Nandeesha, P.; Kaashyap, M.; Kumar, S.; Datta, S. *Indian J. Agric. Biochem.* **2014**, *27*, 81−84.

(34) Tsai, L.-C.; Hsieh, H.-M.; Huang, L.-H.; Wang, J.-C.; Linacre, A.; Lee, J. C.-I. *Forensic Sci. Int.* **2006**, *158*, 250−251.

(35) Wang, P.; Xia, H.; Zhang, Y.; Zhao, S.; Zhao, C.; Hou, L.; Li, C.; Li, A.; Ma, C.; Wang, X. *BMC Genomics* **2015**, *16*, 21.

(36) Tar'an, B.; Warkentin, T.; Somers, D. J.; Miranda, D.; Vandenberg, A.; Blade, S.; Bing, D. *Euphytica* **2004**, *136*, 297−306.

(37) Hashemi, S. H.; Mirmohammadi-Maibody, S. A. M.; Nematzadeh, G. A.; Arzani, A. *Afr. J. Biotechnol.* **2009**, *8*, 2094−2101.

(38) Jordan, M. C.; Somers, D. J.; Banks, T. W. *Plant Biotechnol. J.* **2007**, *5*, 442−453.

(39) Kato, S.; Sayama, T.; Fujii, K.; Yumoto, S.; Kono, Y.; Hwang, T.-Y.; Kikuchi, A.; Takada, Y.; Tanaka, Y.; Shiraiwa, T.; Ishimoto, M. *Theor. Appl. Genet.* **2014**, *127*, 1365−1374.

(40) McCarthy, P. L.; Hansen, J. L.; Zemetra, R. S.; Berger, P. H. *Biotechniques* **2002**, *32* (560), 562−564.

(41) Yin, G.; Xu, H.; Liu, J.; Gao, C.; Sun, J.; Yan, Y.; Hu, Y. *Gene* **2014**, *546*, 177−186.

(42) Shirley, N.; Allgeier, L.; LaNier, T.; Miller Coyle, H. *J. Forensic Sci.* **2013**, *58*, S176−S182.

(43) Lesiak, A. D.; Musah, R. A.; Cody, R. B.; Domin, M. A.; Dane, A. J.; Shepard, J. R. *Analyst* **2013**, *138*, 3424−3432.

(44) Musah, R.; Domin, M.; Cody, R.; Lesiak, A.; Dane, A. J.; Shepard, J. *Rapid Commun. Mass Spectrom.* **2012**, *26*, 2335−2342.

(45) Amokaha, R. A.; Ubwa, S. T.; Otokpa, M.; Shenge, G. *J. Chem. Soc. Nigeria* **2002**, *27* (1), 105−106.

(46) El Bazaoui, A.; Stambouli, H.; Bellimam, M. A.; Soulaymani, A. *Ann. Toxicol. Anal.* **2009**, *21*, 183−188.

(47) Philipof, S.; Berkhov, S. *Z. Naturforsch.* **2002**, *57*, 559−561.

(48) Jakabova, S.; Vincze, L.; Farkas, A.; Kilar, F.; Boros, B.; Felinger, A. *J. Chromatogr. A* **2012**, *1232*, 295−301.

(49) Kim, S. W.; Kim, H. J.; Kim, J. H.; Kwon, Y. K.; Ahn, M. S.; Jang, Y. P.; Liu, J. R. *Plant Methods* **2011**, *7*, 14.

(50) Tata, A.; Perez, C.; Hamid, T.; Bayfield, M.; Ifa, D. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 641−648.

(51) Thunig, J.; Hansen, S. H.; Janfelt, C. *Anal. Chem.* **2011**, *83*, 3256−3259.

(52) Wiseman, J. M.; Ifa, D. R.; Zhu, Y.; Kissinger, C. B.; Manicke, N. E.; Kissinger, P. T.; Cooks, R. G. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 18120−18125.