analytical chemistry

Article

pubs.acs.org/ac

Differentiation of Eight Phenotypes and Discovery of Potential Biomarkers for a Single Plant Organ Class Using Laser Electrospray Mass Spectrometry and Multivariate Statistical Analysis

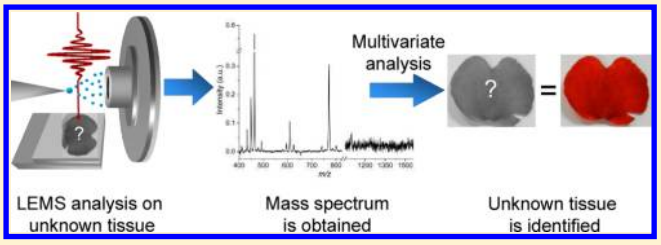
Paul M. Flanigan, IV,[†] Laine L. Radell,[†] John J. Brady,[‡] and Robert J. Levis*,[†]

[†]Department of Chemistry, Temple University, 1901 N. 13th St., Philadelphia, Pennsylvania 19122, United States

[‡]U.S. Army Research Laboratory, 2800 Powder Mill Road, Adelphi, Maryland 20783, United States

Supporting Information

ABSTRACT: Laser electrospray mass spectrometry (LEMS) coupled with offline multivariate statistical analysis is used to discriminate eight phenotypes from a single plant organ class and to find potential biomarkers. Direct analysis of the molecules from the flower petal is enabled by interfacing intense (10^{13} W/cm^2), nonresonant, femtosecond laser vaporization at ambient pressure with electrospray ionization for postionization of the vaporized analytes. The observed mass spectral signatures allowed for the discrimination of various phenotypes using principal component analysis (PCA) and either linear discriminant analysis (LDA) or K-nearest neighbor (KNN) classifiers. Cross-validation was performed using multiple training sets to evaluate the predictive ability of the classifiers, which showed 93.7% and 96.8% overall accuracies for the LDA and KNN classifiers, respectively. Linear combinations of significant mass spectral features were extracted from the PCA loading plots, demonstrating the capability to discover potential biomarkers from the direct analysis of tissue samples.



Mass spectral analysis techniques, such as gas chromatography/mass spectrometry (GC/MS)[1] and liquid chromatography−mass spectrometry[2] often require extensive sample preparation for analysis and identification of molecules from tissue samples including, for example, homogenization, filtration, and extraction. Direct analysis can be performed on whole tissue samples to provide spatial information for imaging profiles.[3−5] Secondary ion mass spectrometry (SIMS)[6−8] and matrix-assisted laser desorption ionization (MALDI)[3−6,9−12] have been extensively used for ex vivo tissue analysis; however, the tissue sample has to be freeze-dried prior to analysis due to the high vacuum experimental conditions. Atmospheric pressure mass spectral techniques, such as desorption electrospray ionization (DESI),[13−18] leaf spray,[19] atmospheric pressure MALDI (AP-MALDI),[20] electrospray-assisted laser desorption ionization (ELDI),[21] and laser ablation electrospray ionization (LAESI),[22−25] permit ex vivo analysis of the tissue sample in its native state. For example, DESI uses an electrospray plume to impinge a tissue sample resulting in desorption of molecules through a droplet pickup mechanism. Leaf spray, derived from paper spray,[26] utilizes the plant itself as an electrospray source when an electrical potential is applied to the leaf. Although leaf spray allows for direct mass analysis of certain molecules from plant samples, no spatial information can be obtained. DESI and laser-based methods, such as AP-MALDI, ELDI, and LAESI have enabled spatially resolved mass analysis of tissues, usually on the order of ∼200 μm. Recently, nano-DESI[27] and atmospheric-pressure scanning microprobe MALDI[28] have achieved higher spatial resolutions of ∼10 μm for tissue imaging. In addition to lateral spatial information, laser-based methods allow for depth profiling.[23,24,29]

Recently, direct tissue analysis was enabled by a novel laser-based mass spectrometric technique called laser electrospray mass spectrometry (LEMS). Different *Impatiens* plant organs (flower petal, leaf, and stem) and green and white regions of the zebra plant were mass analyzed using LEMS and discriminated with an offline classifier.[29] LEMS is an ambient mass spectrometric method that utilizes a nonresonant, femtosecond (fs) laser for vaporization prior to postionization by an ESI source. The fs laser pulse vaporizes the nonvolatile sample material without the necessity of a particular resonant transition in the sample or a matrix due to the nonlinear absorption of the laser pulse. The nonresonant absorption of the radiation is enabled by the high intensity of the laser pulse (10^{13} W/cm^2). The gas phase analyte released as a result of the process is subsequently captured and ionized by an electrospray plume prior to transfer into the mass spectrometer for mass analysis. In this manner, LEMS has enabled mass spectral analysis for small biomolecules,[30] proteins,[31,32] lipids,[33] explosives,[34,35] narcotics, and pharmaceuticals[36] adsorbed onto surfaces at atmospheric pressure, without extensive sample preparation.

Received: May 7, 2012
Accepted: June 13, 2012
Published: June 13, 2012

   dx.doi.org/10.1021/ac3012335 | *Anal. Chem.* 2012, 84, 6225−6232

The application of mass spectrometry to the analysis of complex mixtures often requires multidimensional or multivariate statistical analysis for the characterization and differentiation of the constituents composing the sample. Multivariate data analysis reveals the relationship between many variables, and in the case of mass spectrometry, these consist of the distribution of mass-to-charge (m/z) ratios. These relationships can then be used for the detection of similar profiles, the differentiation between profiles, and classification of unknown mass spectra according to predefined classes.[37] Statistical analysis has been previously used to classify mass spectra of tissue samples obtained with GC/MS,[1] SIMS,[8] and MALDI.[11,12] Hierarchical cluster analysis and principal component analysis (PCA) were used for phenotype classification of *Arabidopsis* leaves based on GC/MS metabolic fingerprinting.[1] Individual cell types were identified within heterogeneous cultures through the use of PCA and partial least-squares-discriminant analysis (PLS-DA) of SIMS mass spectra.[8] In addition, PCA was used to differentiate regions of animal tissues in mass spectral images obtained with MALDI.[11,12]

Multivariate analysis frequently starts with a dimensionality reduction method, such as PCA, to address the problem of high dimensionality and small sample sizes, thus reducing the size of the variable set prior to analysis.[38] PCA transforms raw variables (i.e., m/z ratios for mass spectral data) from the data set into a limited number of new variables, called principal components (PCs), that are constructed as linear combinations of the original variables.[37,39,40] Principal components are orthogonal functions that contain the maximum amount of variance in the data set with a minimum number of functions. Principal components are often ordered according to the decreasing amount of variance (e.g., PC1 contains the largest amount of variance; PC2, orthogonal to PC1, displays the next largest amount of variance; etc.). Extracting a handful of variables from a large initial data set not only simplifies the classification but also often results in the discovery of biomarkers, and this is one motivation for performing biological mass spectrometry experiments,[38] especially on tissues. For example, mass spectrometry combined with multivariate discriminant analysis and classification procedures can lead to the differentiation of tissue samples into classes, such as diseased vs nondiseased. The biomarkers characteristic to each class could be used for diagnostic purposes.[41]

Two previous LEMS investigations were subjected to multivariate data analysis that resulted in high fidelity classifications. The first investigation involved the discrimination of different plant tissues using a compressive linear classifier (CLC) and linear discriminant analysis (LDA).[29] Briefly, the CLC reduced the dimensions of the raw mass spectral data with an empirically set intensity threshold. A matrix (M) containing the reduced mass spectra was multiplied by a randomized matrix (T), yielding a transformed matrix (M_T) that was classified using LDA. The process of applying random perturbations to the T matrix and reclassification of M_T was repeated iteratively until misclassifications in the LDA process were minimized. The CLC resulted in 98.5% overall accuracy and outperformed principal component regression (PCR) analysis for this data set. The second investigation concerned the discrimination of improvised explosive device (IED) signatures for inorganic salts.[35] Forty characteristic features were used to create a library and were integrated for all mass spectra. After PCA was performed on the training set, the resultant eigenvector matrix was multiplied by the mean-adjusted testing set matrix. Classification of the inorganic IEDs performed by LDA resulted in 99.0% accuracy.

Here, we further explore the use of LEMS with an offline classifier, a combination of PCA with LDA or K-nearest neighbor (KNN) classifiers, to discriminate various phenotypes of the same organ type. We demonstrate that the mass spectra of the various phenotypes of *Impatiens* plant flower petals give different mass spectral features, and these can be used to categorize the phenotype with a high success rate. Linear combinations of the mass spectral features responsible for phenotype differentiation are identified as potential biomarkers through the use of PCA loading plots.

EXPERIMENTAL SECTION

Sample Preparation. Eight phenotypes (Figure S1, Supporting Information; white, red, hot pink, light pink, light purple, purple, orange, and coral) of the *Impatiens* plant flower petals were obtained from the Temple University campus. The flower petals were cut in 5 mm × 15 mm rectangles immediately before analysis and affixed to a stainless steel slide using double-sided tape. The sample slide was then placed in the LEMS source chamber on a metal plate that was situated on a three-dimensional translational stage, which allowed for the stage and the flower petal to be raster scanned in order for new sample to be analyzed with every laser pulse.

Laser Vaporization. The laser vaporization, ionization, and detection apparatus has been previously described in detail.[29−36] Briefly, the laser system used for vaporization is composed of an oscillator (KM Laboratories, Inc., Boulder, CO, USA) that seeds a regenerative amplifier (Coherent, Inc., Santa Clara, CA, USA), which creates 70 fs, 2.5 mJ pulses centered at 800 nm. The laser's repetition rate was set to 10 Hz to allow for fresh sample to be vaporized for each laser pulse and to enable synchronization with the hexapole of the electrospray ion source,[30] which was operated in trapping mode. The fs laser pulse was focused to a spot size of ∼300 μm in diameter using a 16.9 cm focal length lens, with an incident angle of 45° with respect to the sample. The intensity of the laser at the area sampled was approximately 2×10^{13} W/cm^2. The vaporization volume is approximately 300 μm in diameter and 10 μm deep. The sample was positioned 6.4 mm below and 1 mm in front of the electrospray needle. The steel sample plate was biased to −2.0 kV to compensate for the distortion of the electric field between the capillary and the needle caused by the introduction of the sample stage. As a result of laser irradiation, analyte was ejected in a direction perpendicular to the electrospray plume, where capture and ionization occurred.

Ionization and Mass Spectrometry. The electrospray source (Analytica of Branford, Inc., Branford, CT, USA) used to capture, ionize, and transfer the vaporized material utilizes an electrospray needle, dielectric capillary, skimmer, and a hexapole. The electrospray source was operated in positive ion mode wherein the electrospray needle was maintained at ground while the inlet capillary was biased to −4.5 kV. The ESI needle was 6.4 mm in front of the capillary entrance and 6.4 mm above and parallel to the sample stage. The acidified electrospray solvent, 1:1 (v:v) water−methanol (Fisher Scientific, Fair Lawn, NJ, USA) with 1% acetic acid (Fisher Scientific), was pumped through the needle by a syringe pump (Harvard Apparatus, Holliston, MA, USA) at a flow rate of 3 μL/min. The electrospray plume was dried by counter current nitrogen gas at 180 °C before entering the inlet capillary.

dx.doi.org/10.1021/ac3012335 | *Anal. Chem.* 2012, 84, 6225−6232

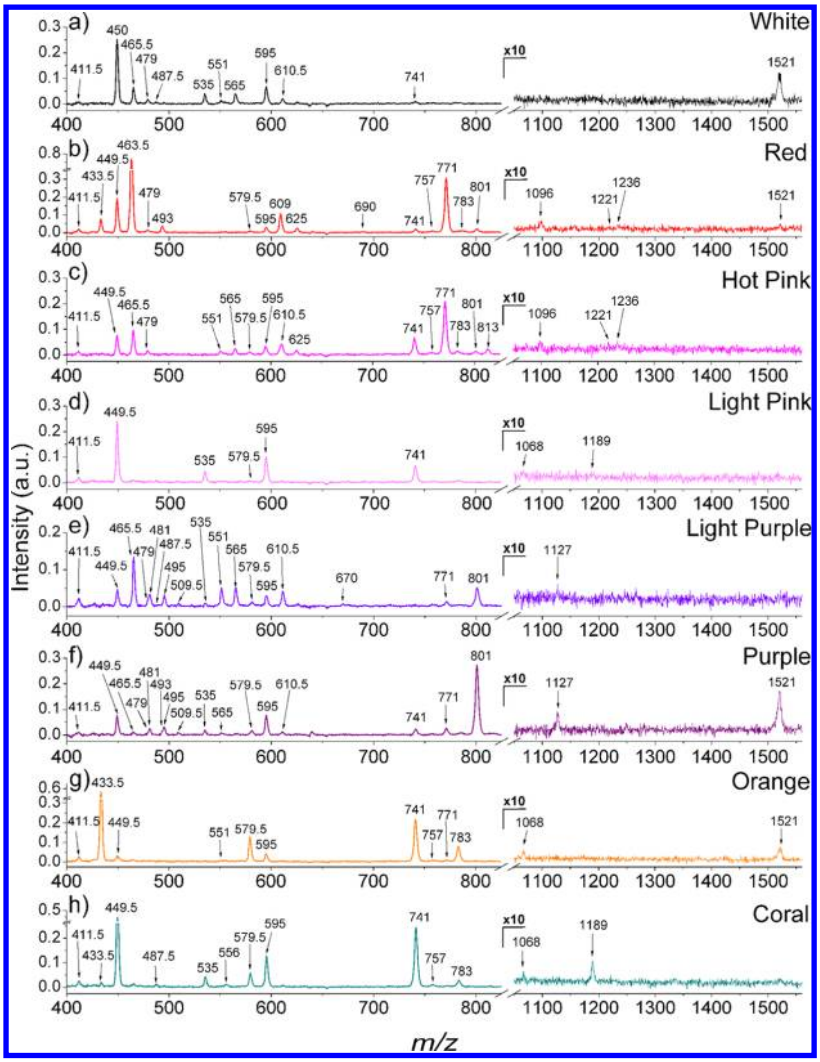


Figure 1. LEMS mass spectra (blank-subtracted) of the higher m/z region for the various *Impatiens* plant flower petals labeled with their respective phenotypes. The features above m/z 1050 are magnified 10×. Peaks, which are identified with their m/z values, are listed in Table S2 (Supporting Information) with their possible corresponding biological compounds.

The created ions were transferred to a hexapole contained within the electrospray ion source via a dielectric capillary. Trapping was performed in the hexapole for 250 μs at a repetition rate of 10 Hz to increase the ion yield. After trapping, the ions were transferred into a pulsed orthogonal time-of-flight (o-TOF), via a RF only hexapole (transmission mode), where mass separation and detection occurred. Each experiment consisted of averaging 50 laser shots (10 Hz repetition rate for a total of 5 s acquisition time) for a mass spectrum, which were averaged using a digital oscilloscope (LeCroy Wavesurfer 422, LeCroy Co., Chestnut Ridge, NY, USA). Three separate flower petals of each phenotype were analyzed with 10 experiments performed on each sample for a total of 30 experiments of each phenotype.

An ESI solvent background mass spectrum was acquired before vaporization of each sample set to allow background subtraction of solvent-related peaks. Negative features observed in the blank-subtracted mass spectra result from the vaporized analytes competing for charge and thus altering the solvent ion distribution. Positive features in the mass spectrum that are not labeled in the blank-subtracted mass spectra are solvent-related features.

The ESI-o-TOF mass spectrometer, constructed in-house, was calibrated using an ESI tuning solution (Agilent Technologies, Inc., Santa Clara, CA, USA) containing species that form ions with known m/z values from m/z 100 to 3000. The flight times of the ionized species were used to calibrate the m/z range of the ESI-o-TOF using a quadratic fit in a custom Labview 8.5 program (National Instruments, Austin, TX, USA). The estimated error for m/z values, which was calculated using the theoretical m/z values and the experimental m/z values for the analytes from the ESI tuning solution, was approximately 0.1% (1000 ppm). The resolution of the instrument was determined to be ~220 m/Δm using the fwhm for the peak at m/z 463.5 from the red flower phenotype. Note that, due to the mass accuracy and the low resolution of the instrument, the experimental m/z values are rounded to the nearest 0.5 m/z and thus may differ slightly from reported values.

Principal Component Analysis and Classification. The raw mass spectral data files obtained from the digital oscilloscope were saved using a custom Labview 8.5 program and imported into Origin 7.5 (OriginLab, Northampton, MA, USA) for data manipulation. The major preprocessing steps prior to classification included baseline subtraction, which was

dx.doi.org/10.1021/ac3012335 | *Anal. Chem.* 2012, 84, 6225−6232

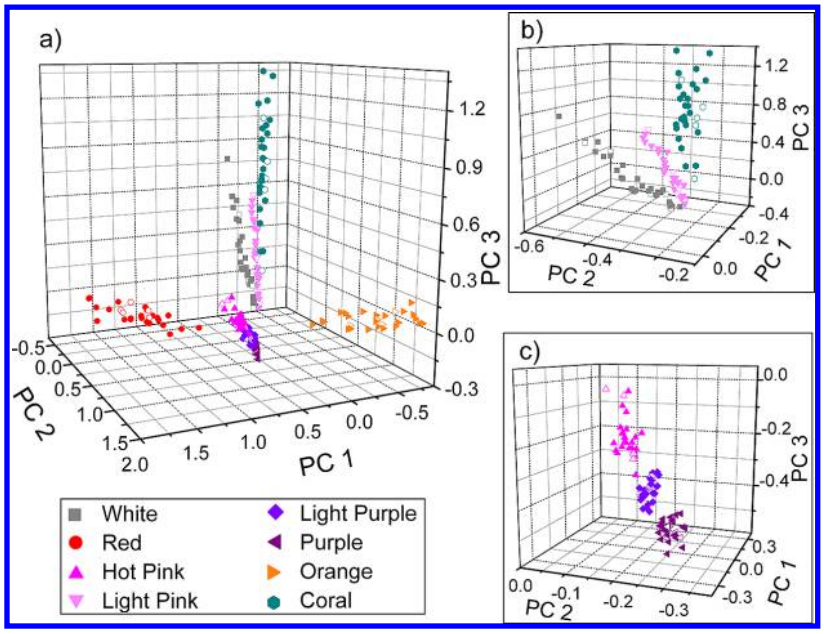


Figure 2. (a) PCA analysis projected into three of the six dimensions for white (square), red (circle), hot pink (triangle), light pink (triangle pointing down), light purple (diamond), purple (triangle pointing left), orange (triangle pointing right), and coral (hexagon) Impatiens plant flower petals. The open and the filled colored symbols represent the training and testing sets for each phenotype, respectively. The PCA plots are magnified for (b) the white, light pink, and coral flower petals and for (c) the pink, light purple, and purple flower petals to show separation among the groups.

written into the Labview code used to save the mass spectra, and spectral realignment to a solvent feature common to all the mass spectra at m/z 279.0 (not shown in blank-subtracted mass spectra in Figure 1).

From the 30 experiments obtained for the eight phenotypes, a randomized training set was created consisting of 5 spectra for each of the eight phenotypes, yielding a 40 × 4300 matrix. Note that the m/z range was reduced from m/z 1 to ~5600 (10 000 data points) to m/z ~75 to 1600 (4300 data points) for quicker computations. PCA was performed on the training set using Matlab, yielding a transformed data set [mean-adjusted training set (40 × 4300 matrix) times the calculated eigenvectors (4300 × 4300 matrix)]. The eigenvectors, also known as coefficients or loadings, obtained from PCA of the training set were then multiplied by the mean-adjusted data of the testing set, which contained the remaining 25 spectra of each of the eight phenotypes (resulting in a 200 × 4300 matrix). Classification of the testing data was performed in Matlab using linear discriminant analysis or K-nearest neighbor (k = 2, cityblock metric) classifiers.

Cross-validation (CV) methods have been used to evaluate the predictive ability of statistical analysis methods in the classification of mass spectral data.[38,42−46] To assess the overall accuracy and robustness of our predictive model, repeated subsampling cross-validation was performed using the phenotype tissue LEMS data. The 30 mass spectra for each tissue phenotype (240 total) were split into six different sets, where the training sets consisted of 5 random spectra for each phenotype (40 total) and the testing sets composed of the remaining 25 spectra for each phenotype (200 total). After a mass spectrum was used in a training subset, it was ruled out for other training subsets, so each spectrum of the 240 total spectra was used in a training set exactly once. Table S1 in the Supporting Information lists the mass spectra of each phenotype used for the CV training sets. Low values for the root-mean-square deviations (rmsd) of the cross-validation reveals robustness in the classification procedure.[39] The rmsd metric arises from analysis of the spread of classification error as the training set cycles through the data set.

Safety Considerations. Appropriate laser eye protection was worn by all lab personnel.

RESULTS AND DISCUSSION

Laser Electrospray Mass Spectrometry of Impatiens Plant Flower Petals. In a previous study, nonresonant fs laser vaporization with electrospray postionization was successfully used as a means for differentiation of diverse plant organ types (*Impatiens* leaf, stem, and petal; green and white zebra plant leaf).[29] In the current investigation, LEMS and multivariate statistical methods were used to discriminate eight different phenotypes (Figure S1, Supporting Information) of *Impatiens* plant flower from the same organ class. Figure S2 (Supporting Information) shows the representative LEMS mass spectra of the flower petals in the lower mass region. Most of the peaks below m/z 400, such as those at m/z 165.0, 177.0, 194.5, 198.5, 219.0, and 344.5, are common to all phenotypes. Some of these features correlate with the small organic acids *p*-coumaric acid, ferulic acid−H_2O, and ferulic acid at m/z 165.0, 177.0, and 194.5, respectively,[47,48] which are typically found as acylations of major compounds found in plants. Although these phenolic moieties are usually observed in negative ion mode ESI due to higher ion abundance, they can also be observed in positive ion mode ESI.[49,50] Table S2 in Supporting Information lists the possible peak assignments based on those found in the literature[47,48,51−57] since tandem MS analysis cannot be performed with the current instrument. Although the majority of the features are common among the mass spectra, some peaks may help distinguish the different phenotypes, such as the unique features observed at m/z 284.0, 287.5, and 301.0 for the light purple (Figure S2e, Supporting Information), coral (Figure S2h, Supporting Information), and red (Figure S2b, Supporting Information) mass spectra, respectively. Two of

dx.doi.org/10.1021/ac3012335 | *Anal. Chem.* 2012, 84, 6225−6232

these features likely correspond to cyanidin (m/z 287.5) and peonidin (301.0) which are anthocyanidins, known plant pigments that contribute to coloring of the coral and red flower petals, respectively.[58,59]

The representative LEMS mass spectra of the higher m/z region (m/z 400−850) for the eight phenotypes (Figure 1) contain the peaks with the highest ion abundances in the mass spectra for all phenotypes. The major features observed in this region, m/z 433.5, 449.5, 463.5, 465.5, 535.0, 551.0, 565.0, 579.5.0, 595.0, 609.0, 610.5, 741.0, 771.0, 783.0, and 801.0 are most likely due to flavonoids, including water-soluble anthocyanins that constitute the different phenotypes of the flower petals. Since few peaks are common to all phenotypes, e.g., m/z 449.5 and 595.0, this mass spectral region should contain distinguishing features. Some peaks in the mass spectra are common to all phenotypes except for one phenotype, such as m/z 579.5 and 741.0 that are absent in the white (Figure 1a) and light purple (Figure 1e) mass spectra, respectively. Unique features observed in the mass spectra for individual phenotypes include m/z 463.5, 609.0, and 690.0 for the red phenotype (Figure 1b), 813.0 for the hot pink phenotype (Figure 1c), 670.0 for the light purple phenotype (Figure 1e), 481.0, 495.0, and 509.5 for the purple phenotypes [purple (Figure 1f) and light purple (Figure 1e)], and 556.0 for the coral phenotype (Figure 1h). Distinguishing the eight phenotypes without multivariate analysis would be an arduous task as there are more than 50 relevant features in the mass spectra.

Multivariate Classification. In this current investigation, LEMS mass spectra obtained from the eight *Impatiens* phenotypes, as seen in Figure 1, were subjected to multivariate data analysis to evaluate the potential for discrimination of different phenotypes from the same organ type. Unlike the previous classifications of LEMS data,[29,35] the samples were of the same organ type and the classification was performed on the whole mass spectral range of interest rather than focusing on the integrated intensities of selected signature features as was performed in a prior investigation.[35] Figure 2a shows the projection of the training and testing data from CV set 1 using the first three PCs. The training set and testing set are denoted by the unfilled and filled symbols, respectively. The PCA projection plots are magnified to show separation among white, light pink and coral flower petals in Figure 2b and hot pink, light purple, and purple flower petals in Figure 2c. As can been seen, the PCA procedure clustered the different phenotypes well, allowing for the accurate categorization with LDA and KNN classifiers.

To extract the maximum information from a mass spectrometric data set, one needs to make certain assumptions regarding data generation. Linear discriminant analysis, a parametric approach, assumes that the data has a Gaussian probability distribution[38] and is one of the most widely used classification methods for mass spectrometry data. Because LDA struggles with the high dimensionality problem, a dimensionality reduction method, such as PCA[35,60] or its supervised complement, PLS,[44] is usually performed before LDA to ensure the best classification results.[39] The K-nearest neighbor classifier is nonparametric and makes no prior assumptions about the training data.[38] KNN classifies an object to the class most common from the training set in the local vicinity, which is defined by the parameter k, the number of neighbors to be used for classification. For example, if k = 3, an object is classified on the basis of the majority of the three closest training points and their respective classes.

The results from the classification using LDA and KNN are shown as divisions of the six different CV testing subsets and by flower petal color in Tables 1 and 2, respectively. Although

Table 1. Classification of the Eight *Impatiens* Phenotypes Using PCA and Either LDA or KNN Classifiers for the 6 CV Sets

	linear discriminant analysis		K-nearest neighbor	
CV set	samples correctly characterized	percent (%)	samples correctly characterized	percent (%)
set 1	195/200	97.5	198/200	99.0
set 2	194/200	92.5	194/200	97.0
set 3	189/200	91.0	189/200	94.5
set 4	194/200	91.5	194/200	97.0
set 5	191/200	94.0	191/200	95.5
set 6	195/200	95.5	195/200	97.5
total	1124/1200	93.7	1161/1200	96.8
	root-mean-square deviation	2.3	root-mean-square deviation	1.4

Table 2. Classification of the Eight *Impatiens* Phenotypes Using PCA and Either LDA or KNN Classifiers for Each Flower Petal Color Shown as a Summation of the 6 CV Sets

	linear discriminant analysis		K-nearest neighbor	
flower petal color	samples correctly characterized	percent (%)	samples correctly characterized	percent (%)
white	118/150	78.7	131/150	87.3
red	147/150	98.0	150/150	100.0
hot pink	150/150	100.0	149/150	99.3
light pink	135/150	90.0	148/150	98.7
light purple	150/150	100.0	150/150	100.0
purple	150/150	100.0	150/150	100.0
orange	146/150	97.3	150/150	100.0
coral	128/150	85.3	133/150	88.7
total	1124/1200	93.7	1161/1200	96.8
	root-mean-square deviation	7.6	root-mean-square deviation	5.1

KNN slightly outperformed LDA, both methods classified the testing sets with high fidelity, resulting in accuracies of 96.8% and 93.7%, respectively. The rmsd of the cross-validation were 2.3% and 1.4% when comparing among CV sets (Table 1) and 7.6% and 5.1% when comparing among phenotypes (Table 2) for LDA and KNN, respectively. The rmsd values were higher for the phenotype data as there was more variation in the range of correctly classified spectra among the flower petals (i.e., 78.7% to 100.0% for LDA) than the CV sets (i.e., 91.0% to 97.5% for LDA). These low rmsd values show the robustness of the classification procedure for the current data set. The phenotypes with the highest percentage of misclassifications, in decreasing order, were the white, coral, and light pink flower petals. This is not surprising given the fact that the three phenotypes are closely grouped together in the three-dimensional PCA projection plot (Figure 2b). Note that only 77.1% of the variance was explained by the first three PCs and it takes nine PCs to explain 95.7% of the variance (as seen in Figure S3, Supporting Information). The distribution of variance into nine components offers an explanation for the minor number of misclassifications using the first three components. Variations in the relative peak intensities are

dx.doi.org/10.1021/ac3012335 | Anal. Chem. 2012, 84, 6225−6232

expected since flower petals may not be in the same stage of growth and may have varying amounts of the same compound,[61] which may also affect the classification, as LDA has a Gaussian distribution bias. However, the clear separation in three-dimensional space and the accurate classification of the different phenotypes, as evaluated with cross-validation, demonstrates that PCA with LDA or KNN classifiers can be can be used to discriminate a large number of variations of the same organ type.

The fact that multivariate analysis of the 50 relevant features in the mass spectra can be used to distinguish the eight phenotypes suggests that the reproducibility of the LEMS measurement is good. A previous LEMS investigation[29] of different plant tissues (petal, stem, and leaf) showed good reproducibility for the mass spectral response for measurements on different locations of the same tissue sample and on different tissue samples. We have shown that the LEMS signal scales linearly with concentration in a two component lipid system.[33] Such scaling is also beneficial for multivariate analysis. In addition, it should be noted that the limit of detection for LEMS has been shown to be about 2 orders of magnitude less sensitive than conventional ESI-MS,[33] which is due to the ~1% neutral capture efficiency for the current apparatus.[36]

Biomarker Discovery. For biological tissue analysis, the elucidation of biomarkers is important for the discovery of indicators for disease diagnosis. In closely related tissue types, no single mass feature is likely to be a biomarker, rather the linear combination of multiple masses will form the biomarker. This is precisely what PCA represents. To determine if the current classification method was capable of discovering biologically relevant features in high dimensional space from the LEMS phenotype plant flower petal data, the PCA loadings (for training subset 1) were plotted for the first three PCs in Figure 3. The loading value of a variable (i.e., a particular m/z ratio for the mass spectral data) indicates how much of that variable contributes to defining the PC.[40] The m/z ratios that contribute the most to the PC have loading values furthest from zero. The set of PCs calculated from the mass spectral training set diagonalizes the matrix of mass spectral features, thus providing the maximal separation for each phenotype comprising the training set, ideally in reduced dimension. The use of PCA is necessary when each phenotype provides a distinct linear combination of the set of mass spectral features. For a given PC, variables correlated to a phenotype have loading values with similar signs (either positive or negative) to phenotypic score values whereas anticorrelated variables have opposite signs. For example, the variables m/z 433.5 and 741.0 have negative PC1 loadings and are positively correlated with the orange phenotype, which is projected into the negative PC1 scores (Figure 2). On the other hand, the variables m/z 463.5, 609.0, and 771.0 have positive PC1 loadings and thus are negatively correlated to the orange phenotype. As seen in Figure 1g, the features at m/z 433.5 and 741.0 are present in the orange phenotype mass spectrum while the features at m/z 463.5, 609.0, and 771.0 are either not observed or have very low ion abundance. Thus, the potential biomarkers for discrimination of *Impatiens* plant flower petals are composed of the PCs derived primarily from the features above m/z 400 in the mass spectra, specifically linear combinations of m/z 433.5, 449.5, 463.5, 579.5, 595.0, 609.0, 741.0, 771.0, 783.0, and 801.0. Chemically, these features are mostly variants of anthocyanins (based on probable assignments (Table S2, Supporting Information)), including cyanidin, malvidin, pelargonidin,

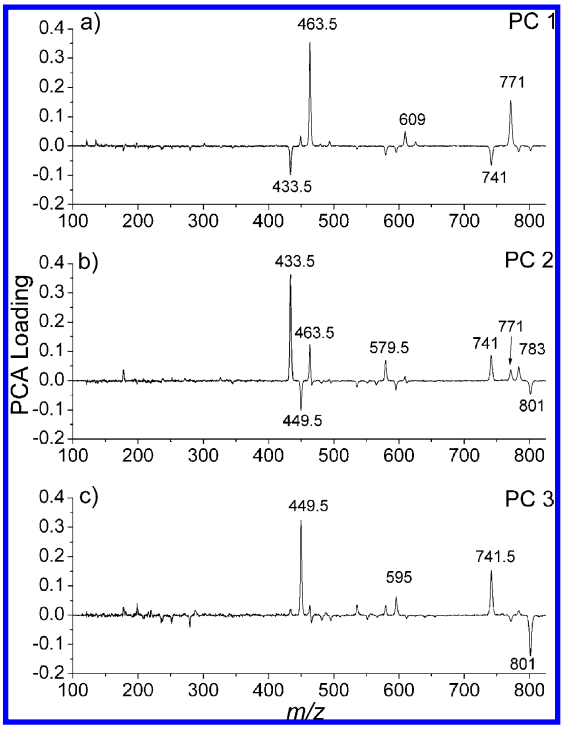


Figure 3. PCA loading plots for the first three PCs, (a) through (c), respectively, from the LEMS analysis of the *Impatiens* plant flower petals that show important mass spectral features, which are labeled with m/z values for clarification, that comprise the potential biomarkers. The plots are from multivariate statistical analysis performed on training set 1 from the cross-validation study.

peonidin, and petunidin, which give rise to the different colors of the flower petals that define their respective phenotypes.[58] The PCA loading results of the first three principal components showed the same set of mass features for all six cross-validation training sets (Figure S4, Supporting Information). The ability to extract important variables from the PCA loading plots demonstrates that multivariate data analysis on LEMS mass spectra can be used to discover potential biomarkers for the discrimination of tissue samples.

■ CONCLUSION

The ability to classify complex biological samples is an important step toward diagnosing diseased tissues. Discovering the linear combinations of features in high dimensional space that distinguish the classes is an important route to biomarker development for diagnosis because, in complex samples, the probability that a single feature will provide robust classification is small. We have demonstrated that LEMS is capable of direct mass spectral analysis of ex vivo plant tissue samples for discrimination of various phenotypes. The fs laser pulse nonresonantly vaporizes molecules directly from the plant tissue into the gas phase where they are captured by the electrospray solvent, ionized, and mass analyzed. Analysis of the mass spectra of the phenotypes reveals that multiple features (correlated to biologically relevant molecules) are required for discrimination. All eight phenotypes of the *Impatiens* flower petals were differentiated with high accuracy using an offline classifier, consisting of PCA followed by LDA or KNN. The robustness of the classification was confirmed with a cross-validation study. Important variables, which correlate well with compounds relevant to plant flower petals, were extracted from

dx.doi.org/10.1021/ac3012335 | *Anal. Chem.* 2012, 84, 6225−6232

the PCA loading plots, demonstrating the ability to discover potential biomarkers from direct tissue analysis using LEMS. The capability to distinguish eight phenotypes from a single plant organ class is a precursor for the analysis and classification of more complicated biological tissue samples, such as diseased human tissues.

■ ASSOCIATED CONTENT

Supporting Information

Additional information as noted in text. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: rjlevis@temple.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The work was supported by the Office of Naval Research (N00014-10-0293) and the National Science Foundation (CHE 0518497, CHE0957694). J.J.B. acknowledges the Oak Ridge Associated Universities Postdoctoral Research Program.

■ REFERENCES

(1) Messerli, G.; Nia, V. P.; Trevisan, M.; Kolbe, A.; Schauer, N.; Geigenberger, P.; Chen, J.; Davison, A. C.; Fernie, A. R.; Zeeman, S. C. *Plant Physiol.* **2007**, *143*, 1484−1492.

(2) De Vos, R. C. H.; Moco, S.; Lommen, A.; Keurentjes, J. J. B.; Bino, R. J.; Hall, R. D. *Nat. Protoc.* **2007**, *2*, 778−791.

(3) Caldwell, R. L.; Caprioli, R. M. *Mol. Cell. Proteomics* **2005**, *4*, 394−401.

(4) Chaurand, P.; Schwartz, S. A.; Caprioli, R. M. *J. Proteome Res.* **2004**, *3*, 245−252.

(5) Stoeckli, M.; Chaurand, P.; Hallahan, D. E.; Caprioli, R. M. *Nat. Med.* **2001**, *7*, 493−496.

(6) Todd, P. J.; Schaaff, T. G.; Chaurand, P.; Caprioli, R. M. *J. Mass Spectrom.* **2001**, *36*, 355−369.

(7) Brunelle, A.; Touboul, D.; Laprevote, O. *J. Mass Spectrom.* **2005**, *40*, 985−999.

(8) Barnes, C. A.; Brison, J.; Robinson, M.; Graham, D. J.; Castner, D. G.; Ratner, B. D. *Anal. Chem.* **2012**, *84*, 893−900.

(9) Chaurand, P.; Stoeckli, M.; Caprioli, R. M. *Anal. Chem.* **1999**, *71*, 5263−5270.

(10) Caprioli, R. M.; Farmer, T. B.; Gile, J. *Anal. Chem.* **1997**, *69*, 4751−4760.

(11) Trim, P. J.; Atkinson, S. J.; Princivalle, A. P.; Marshall, P. S.; West, A.; Clench, M. R. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 1503−1509.

(12) McCombie, G.; Staab, D.; Stoeckli, M.; Knochenmuss, R. *Anal. Chem.* **2005**, *77*, 6118−6124.

(13) Takats, Z.; Wiseman, J. M.; Gologan, B.; Cooks, R. G. *Science* **2004**, *306*, 471−473.

(14) Wiseman, J. M.; Puolitaival, S. M.; Takats, Z.; Cooks, R. G.; Caprioli, R. M. *Angew. Chem.* **2005**, *117*, 7256−7259.

(15) Wiseman, J. M.; Ifa, D. R.; Song, Q.; Cooks, R. G. *Angew. Chem., Int. Ed.* **2006**, *45*, 7188−7192.

(16) Talaty, N.; Takats, Z.; Cooks, R. G. *Analyst* **2005**, *130*, 1624−1633.

(17) Muller, T.; Oradu, S.; Ifa, D. R.; Cooks, R. G.; Krautler, B. *Anal. Chem.* **2011**, *83*, 5754−5761.

(18) Wu, C.; Ifa, D. R.; Manicke, N. E.; Cooks, R. G. *Analyst* **2010**, *135*, 28−32.

(19) Liu, J.; Wang, H.; Cooks, R. G.; Ouyang, Z. *Anal. Chem.* **2011**, *83*, 7608−7613.

(20) Perdian, D. C.; Schieffer, G. M.; Houk, R. S. *Rapid Commun. Mass Spectrom.* **2010**, *24*, 397−402.

(21) Huang, M. Z.; Hsu, H. J.; Lee, J. Y.; Jeng, J.; Shiea, J. *J. Proteome Res.* **2006**, *5*, 1107−1116.

(22) Nemes, P.; Vertes, A. *Anal. Chem.* **2007**, *79*, 8098−8106.

(23) Nemes, P.; Barton, A. A.; Li, Y.; Vertes, A. *Anal. Chem.* **2008**, *80*, 4575−4582.

(24) Nemes, P.; Barton, A. A.; Vertes, A. *Anal. Chem.* **2009**, *81*, 6668−6675.

(25) Nemes, P.; Woods, A. S.; Vertes, A. *Anal. Chem.* **2010**, *82*, 982−988.

(26) Liu, J.; Wang, H.; Manicke, N. E.; Lin, J. M.; Cooks, R. G.; Ouyang, Z. *Anal. Chem.* **2010**, *82*, 2463−2471.

(27) Laskin, J.; Heath, B. S.; Roach, P. J.; Cazares, L.; Semmes, O. J. *Anal. Chem.* **2012**, *84*, 141−148.

(28) Rompp, A.; Guenther, S.; Schober, Y.; Schulz, O.; Takats, Z.; Kummer, W.; Spengler, B. *Angew. Chem., Int. Ed.* **2010**, *49*, 3834−3838.

(29) Judge, E. J.; Brady, J. J.; Barbano, P. E.; Levis, R. J. *Anal. Chem.* **2011**, *83*, 2145−2151.

(30) Brady, J. J.; Judge, E. J.; Levis, R. J. *Rapid Commun. Mass Spectrom.* **2009**, *23*, 3151−3157.

(31) Judge, E. J.; Brady, J. J.; Levis, R. J. *Anal. Chem.* **2010**, *82*, 10203−10207.

(32) Brady, J. J.; Judge, E. J.; Levis, R. J. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 12217−12222.

(33) Brady, J. J.; Judge, E. J.; Levis, R. J. *J. Am. Soc. Mass. Spectrom.* **2011**, *22*, 762−772.

(34) Brady, J. J.; Judge, E. J.; Levis, R. J. *Rapid Commun. Mass Spectrom.* **2010**, *24*, 1659−1664.

(35) Flanigan, P. M.; Brady, J. J.; Judge, E. J.; Levis, R. J. *Anal. Chem.* **2011**, *83*, 7115−7122.

(36) Judge, E. J.; Brady, J. J.; Dalton, D.; Levis, R. J. *Anal. Chem.* **2010**, *82*, 3231−3238.

(37) Tas, A. C.; Van der Greef, J. *Mass Spectrom. Rev.* **1994**, *13*, 155−181.

(38) Hilario, M.; Kalousis, A.; Pellegrini, C.; Muller, M. *Mass Spectrom. Rev.* **2006**, *25*, 409−449.

(39) Lee, J. L. S.; Gilmore, I. S. In *Surface Analysis - The Principal Techniques*, 2nd ed.; Vickerman, J. C.; Gilmore, I. S., Eds.; John Wiley & Sons, Ltd.: Chichester, UK, 2009, pp 563−612.

(40) Wold, S.; Esbensen, K.; Geladi, P. *Chemom. Intell. Lab.* **1987**, *2*, 37−52.

(41) Norris, J. L.; Cornett, D. S.; Mobley, J. A.; Andersson, M.; Seeley, E. H.; Chaurand, P.; Caprioli, R. M. *Int. J. Mass Spectrom.* **2007**, *260*, 212−221.

(42) Lee, K. R.; Lin, X.; Park, D. C.; Eslava, S. *Proteomics* **2003**, *3*, 1680−1686.

(43) Wagner, M.; Naik, D.; Pothen, A. *Proteomics* **2003**, *3*, 1692−1698.

(44) Purohit, P. V.; Rocke, D. M. *Proteomics* **2003**, *3*, 1699−1703.

(45) Tatay, J. W.; Feng, X.; Sobczak, N.; Jiang, H.; Chen, C. F.; Kirova, R.; Struble, C.; Wang, N. J.; Tonellato, P. J. *Proteomics* **2003**, *3*, 1704−1709.

(46) Villagra, E.; Santos, L. S.; Vaz, B. G.; Eberlin, M. N.; Felipe Laurie, V. *Food Chem.* **2012**, *131*, 692−697.

(47) Giusti, M. M.; Rodriguez-Saona, L. E.; Griffin, D.; Wrolstad, R. E. *J. Agric. Food. Chem.* **1999**, *47*, 4657−4664.

(48) Wu, X.; Prior, R. L. *J. Agric. Food. Chem.* **2005**, *53*, 3101−3113.

(49) Biesaga, M.; Pyrzynska, K. *J. Chromatogr., A* **2009**, *1216*, 6620−6626.

(50) Fang, N.; Yu, S.; Prior, R. L. *J. Agric. Food. Chem.* **2002**, *50*, 3579−3585.

(51) Lin, L. Z.; He, X. G.; Lindenmaier, M.; Yang, J.; Cleary, M.; Qiu, S. X.; Cordell, G. A. *J. Agric. Food. Chem.* **2000**, *48*, 354−365.

(52) Wang, J.; Sporns, P. *J. Agric. Food. Chem.* **1999**, *47*, 2009−2015.

(53) Heier, A.; Blaas, W.; DroB, A.; Wittkowski, R. *Am. J. Enol. Vitic.* **2002**, *53*, 78−86.

(54) Kumar, N.; Bhandari, P.; Singh, B.; Bari, S. S. *Food Chem. Toxicol.* **2009**, *47*, 361−367.
(55) Huang, Z.; Wang, B.; Williams, P.; Pace, R. D. *LWT-Food Sci. Technol.* **2009**, *42*, 819−824.
(56) Wang, H.; Edward, J.; Shrikhande, A. J. *J. Agric. Food. Chem.* **2003**, *51*, 1839−1844.
(57) Li, Y.; Shrestha, B.; Vertes, A. *Anal. Chem.* **2008**, *80*, 407−420.
(58) Asen, S.; Stewart, R. N.; Norris, K. H. *Phytochemistry* **1972**, *11*, 1139−1144.
(59) Goto, T.; Kondo, T. *Angew. Chem., Int. Ed.* **1991**, *30*, 17−33.
(60) Lilien, R. H.; Farid, H.; Donald, B. R. *J. Comput. Biol.* **2003**, *10*, 925−946.
(61) Miles, C. D.; Hagen, C. W. *Plant Physiol.* **1968**, *43*, 1347.