analytical chemistry

Article

Cite This: *Anal. Chem.* 2018, 90, 9206−9217

pubs.acs.org/ac

Direct Analysis in Real Time-Mass Spectrometry and Kohonen Artificial Neural Networks for Species Identification of Larva, Pupa and Adult Life Stages of Carrion Insects

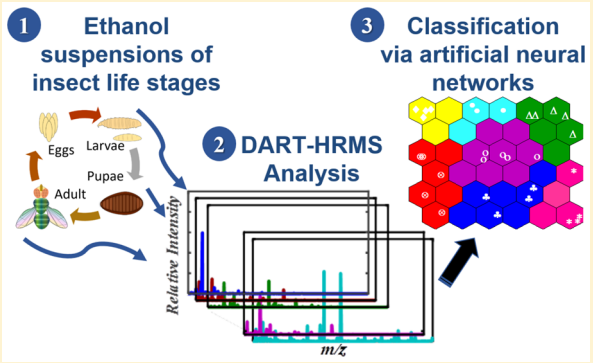
Samira Beyramysoltan,[†] Justine E. Giffen,[†] Jennifer Y. Rosati,[‡] and Rabi A. Musah*[,†]

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

[‡]John Jay College of Criminal Justice, 524 West 59th Street, New York, New York 10019, United States

Supporting Information

ABSTRACT: Species determination of the various life stages of flies (Order: Diptera) is challenging, particularly for the immature forms, because analogous life stages of different species are difficult to differentiate based on morphological features alone. It is demonstrated here that direct analysis in real time-high-resolution mass spectrometry (DART-HRMS) combined with supervised Kohonen Self-Organizing Maps (SOM) enables accomplishment of species-level identification of larva, pupa, and adult life stages of carrion flies. DART-HRMS data for each life stage were acquired from analysis of ethanol suspensions representing *Calliphoridae*, *Phoridae*, and *Sarcophagidae* families, without additional sample preparation. After preprocessing, the data were subjected to a combination of minimum Redundancy Maximal Relevance (mRMR) and Sparse Discriminant Analysis (SDA) methods to select the most significant variables for creating accurate SOM models. The resulting data were divided into training and validation sets and then analyzed by the SOM method to define the proper discrimination models. The 5-fold venetian blind cross-validation misclassification error was below 7% for all life stages, and the validation samples were correctly identified in all cases. The multiclass SOM model also revealed which chemical components were the most significant markers for each species, with several of these being amino acids. The results show that processing of DART-HRMS data using artificial neural networks (ANNs) based on the Kohonen SOM approach enables rapid discrimination and identification of fly species even for the immature life stages. The ANNs can be continuously expanded to include a larger number of species and can be used to screen DART-HRMS data from unknowns to rapidly determine species identity.



Insects of the order Diptera are of high ecological, economic, agricultural, medical, and forensic importance. While they are second only to bees as pollinators, they serve as vectors of a number of serious animal and human diseases. Those of the *Calliphoridae* family are important transmitters of food-borne illnesses and cause millions of dollars in annual losses due to myiasis in livestock. They are also of clinical interest because on one hand they cause myiasis in humans, while on the other, they are used for debridement of nonhealing wounds.[1] Most importantly, they are an essential part of Earth's ecosystem, as they facilitate decomposition of carrion and corpses which they colonize to provide a food source for their larvae. Their association with corpses and carrion can be used to estimate time since death, also known as postmortem interval (PMI), in homicide, animal abuse, and senior/child neglect cases. Adult carrion insects are believed to be attracted to remains by visual and chemical cues that are produced during the course of decomposition, with the corpse serving as a food resource for the larvae that hatch from deposited eggs.[2,3] The blow fly family of flies (Family: Calliphoridae) are of particular utility in this regard because they are usually the "earliest arrivers", often appearing within minutes after a death occurs.[4,5] Since: (1) the predictable order in which different species of insects arrive (termed succession) is directly related to the level of decomposition; and (2) fly development is highly predictable under specific weather and temperature conditions,[6−8] the developmental stage of recovered evidence and knowledge of the species represented can be used to obtain crucial information about the timing of death.[9,10] For this reason, accurate species identification of entomological evidence is an essential aspect of PMI estimations.[11] After oviposition (egg laying), adult flies do not remain near the corpse, and thus immature life stages including eggs, larvae, pupae, and empty puparial casings are the forms of evidence most often collected and identified.[12] Although each stage is readily distinguished

Received: April 16, 2018
Accepted: June 18, 2018
Published: June 20, 2018

DOI: 10.1021/acs.analchem.8b01704
Anal. Chem. 2018, 90, 9206−9217

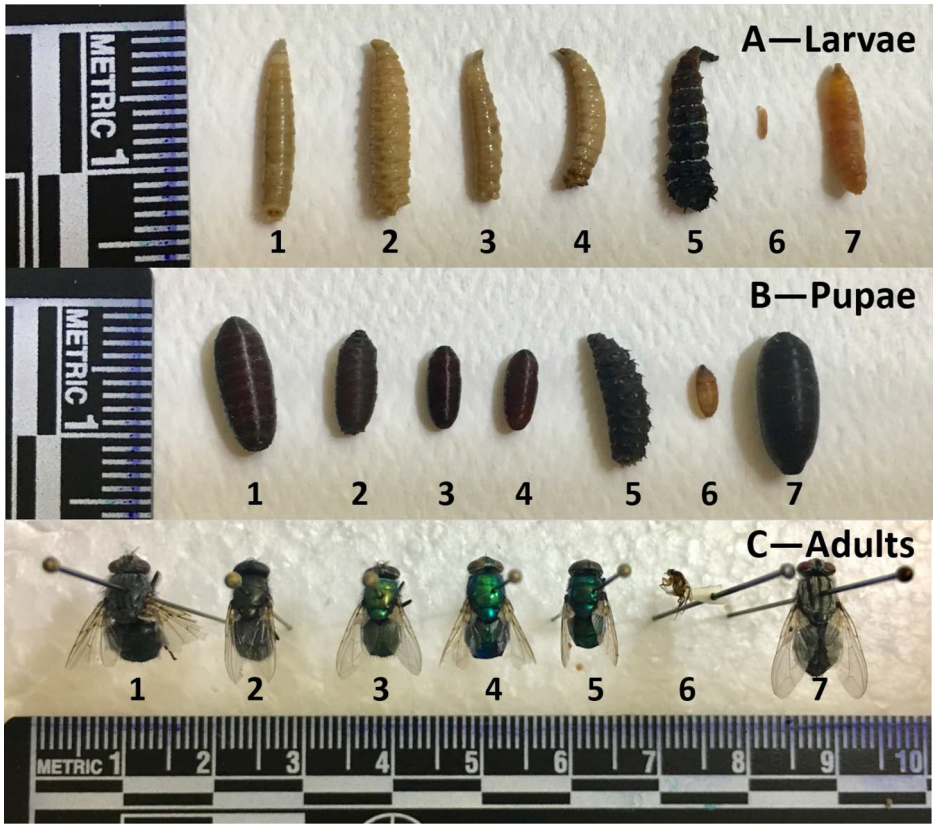


Figure 1. Images of larva (panel A), pupa (panel B), and adult (panel C) life stages of seven blow fly species. In all panels: (1) *C. vicina*, (2) *P. regina*, (3) *L. sericata*, (4) *L. coeruleiviridis*, (5) *C. rufifacies*, (6) *Phoridae* spp., and (7) *Sarcophagidae* spp. Although differentiation between life stages can be readily accomplished by visual examination, different species within the same life stage can be difficult to tell apart. The discoloration of the *C. rufifacies* larva (panel A, 5) is caused by the preservation conditions and is not a reflection of the larva's true color.

from the others, the same life stage of closely related species can be very difficult to differentiate, making identification by visual inspection inaccurate.[13−15] Figure 1 displays the larvae, pupae, and adult life forms of members of the *Calliphoridae* family (*Calliphora vicina* (Robineau-Desvoidy), *Chrysomya rufifacies* (Macquart), *Lucilia coeruleiviridis* (Macquart), *L. sericata* (Meigen), and *Phormia regina* (Meigen)) as well as members of the *Phoridae* and *Sarcophagidae* families. Despite the visually apparent differences between larva (panel A), pupae (panel B), and adults (panel C) being clear to even an untrained eye, the variations in morphology between species within the same life stage can prove challenging. For example, specimen no. 3 (*L. sericata*) and specimen no. 4 (*L. coeruleiviridis*), both members of the *Lucilia* genus, have larvae and pupae that are difficult to distinguish at each stage. For this reason, species identification is commonly accomplished by rearing eggs or larvae into adulthood, in order to utilize gross phenotypic features for visual identification.[16] This practice necessitates the use of entomological keys and is time- and resource-intensive, requiring expert knowledge of insect rearing practices.[17,18] This method is also reliant on discovered eggs, larvae, and/or pupae being viable.

Of the various life stages, species identification by visual inspection is the most difficult for eggs because of their small size and morphological similarity. Thus, they are generally considered to be the least useful form of entomological evidence, despite the crucial information they can provide.[19] Identification of other life stages, such as larvae or pupae, can be accomplished using entomological keys. However, the absence of keys representing a broad range of species limits the utility of this approach.[17,20−22] Furthermore, the results can be highly inaccurate because of the impact of nutritional deficiencies or the presence of drugs in the corpse on the size and weight of larvae and pupae.[23−29] The physical appearance can also be impacted by aging and storage conditions.[23,24,27−30]

Novel species identification methods have been explored as a means to circumvent many of the shortcomings of traditional approaches. Microscope imaging methods, in the form of light and scanning electron microscopy with or without staining, have been used with limited success for the differentiation of blow fly egg genera,[31] but species-level differentiation is generally not possible by this approach.[13,32] DNA typing is another method that has been employed to identify species.[33−36] This method is ideal in situations where an intact or viable insect is not available, because it only requires a small (300 bp) segment of DNA.[37,38] While the technique itself is fairly straightforward, (albeit time- and resource-intensive), its applicability is severely limited by the absence of sequence data for relevant insects.[33,39,40] In addition, sequence variation can differ by less than 25 nucleotides between species, and intraspecific polymorphism sometimes occurs.[37] Nevertheless, identification of DNA markers does have potential applications for determination of geographic distributions of insect populations, which could prove useful in determining if a body was moved posthumously.[41]

More recently, hydrocarbon profiling has been demonstrated as a technique capable of species-level differentiation on a

DOI: 10.1021/acs.analchem.8b01704
Anal. Chem. 2018, 90, 9206−9217

multitude of insect samples.[7,42,43] In a report by Musah et al., the hydrocarbon profiles of empty puparial casings of members of the *Calliphoridae* family were analyzed by direct analysis in real time-high resolution mass spectrometry (DART-HRMS) for the purpose of species identification.[44] Insect cuticular hydrocarbon profiles had been previously established to vary between species,[45−48] and it was shown that the technique could potentially be applied for the determination of insect age. The Musah[44] study demonstrated that multivariate statistical analysis processing of the DART-HRMS data could reveal species identity with a level of statistical certainty. This concept of subjecting DART-HRMS chemical fingerprint profiles to multivariate statistical analysis processing has been applied previously for differentiating plant seeds, identification of endangered wood samples, determination of adulteration of olive oils, identification of lubricants that could be used in sexual assaults, and for plant species identification.[49−57]

While the aforementioned approaches to the analysis of insect life stages were performed on hexane extracts, an analysis method that utilizes aqueous ethanol solutions in which the samples have been immersed would be much more efficient, since this is the form in which the samples are generated in the field and the manner in which they are stored.[27,58] The standard aqueous ethanol solution routinely used for long-term storage of insect evidence provides a medium that can easily and rapidly be analyzed by DART-HRMS to reveal unique and diagnostic chemical fingerprints, including amino acids profiles that are characteristic for a given species.[59] In addition, although previous studies exploited the use of multivariate discriminant analysis techniques to distinguish between species, a statistical analysis method that enables differentiation and identification of a range of species in their different life stages needs to be demonstrated and developed in order for this approach to have practical utility. Furthermore, discriminative variables revealed by the application of appropriate statistical analysis methods would increase the accuracy of species identification. The work reported here illustrates methods for species identification of blow fly larva, pupa, and adult life stages. DART-HRMS spectra of ethanol suspensions of each of these life stages representing *Calliphoridae* species (i.e., *C. vicina, C. rufifacies, L. coeruleiviridis, L. sericata*, and *P. regina*) and members of the *Phoridae* and *Sarcophagidae* families, furnished chemical signatures which, even for closely related species, could be accurately identified by screening them against an artificial neural network that was based on a Kohonen artificial neural network.

METHODS

Chemical Standards. The following chemical standards and solvents were purchased from the indicated vendors: arginine, aspartic acid, glutamine, isoleucine, lysine, proline, threonine, and valine (Sigma-Aldrich, St. Louis, MO); cysteine, glutamic acid, serine, and tyrosine (Oakwood Products, Estill, SC); asparagine, histidine, leucine, and phenylalanine (General Biochemicals, Chagrin Falls, OH); glycine and ethanol (Fisher Scientific, Pittsburgh, PA); alanine (Nutritional Biochemicals Corp., Cleveland, OH); methionine (Mann Research Laboratories Inc., New York, NY); and tryptophan (Calbiochem-Behring Corp., La Jolla, CA).

Collection and Preservation of Necrophagous Fly Samples. Blow fly stock colonies were maintained by Dr. Jennifer Rosati (John Jay College of Criminal Justice, New York, NY). *Phoridae* colonies were started in 2015; *L. sericata* and *Sarcophagidae* colonies were established in 2014; and *C. vicina* and *L. coeruleiviridis* colonies were begun in 2016. These colonies were established with flies gathered from the Manhattan, NY area. *P. regina* and *C. rufifacies* colonies were initially obtained from Dr. Christina Picard at Indiana University-Purdue University (Indianapolis) and Dr. Jason Byrd at University of Florida, respectively, and then established in the entomology lab at John Jay College in 2015. All colonies (except *C. rufifacies*) are annually augmented (until 2017) with wild-type females collected from the Manhattan, NY region via pork liver baited traps.

Adult flies were held in 45 × 45 × 45 cm^3 steel and mesh cages at a temperature of 21 °C with 50% humidity and a diel cycle of 12L:12D. They were fed sugar and milk powder placed in 100 mm plastic Petri dishes (Fisher Scientific, Pittsburgh, PA) and water *ad libitum* in a 500 mL Erlenmeyer flask plugged with paper towels to prevent drowning.

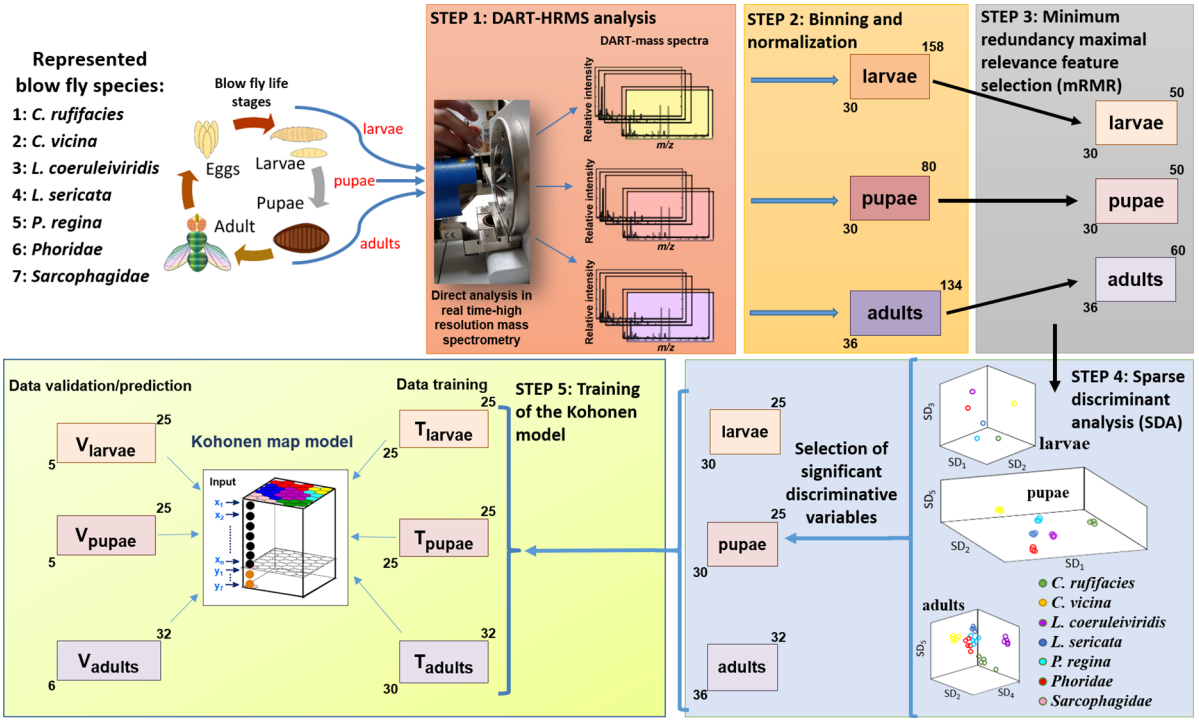
For the collection of entomological samples used in this study, pork liver was used as an oviposition medium and was placed inside a colony cage containing only a single species. Eggs laid on the tissue were subsequently removed and then reared to the larval stage in Mason jars (Ball Corp., Broomfield, CO) with Beta Chip wood shavings (Northeastern Products Corp., Warrensburg, NY) as a pupation medium. The jars were covered with landscape tarp (Scotts Pro Landscape Fabric, Scotts, Marysville, OH) and kept at room temperature. After hatching, larvae were fed pork liver *ad libitum* and kept in rearing jars until adult emergence. At the beginning of each life stage (larva, pupa, and adult) 5−6 individuals of each species were sacrificed by being placed into individual glass vials (Fisher Scientific, Pittsburgh, PA) that were filled with boiling water (a practice used to preserve the tissue), which was then drained and replaced with approximately 2 mL of 70% aqueous ethanol solution (Fisher Scientific, Pittsburgh, PA). These samples were then stored at 4 °C until analysis.

Direct Analysis in Real Time-High-Resolution Mass Spectrometry (DART-HRMS) Experiments. Suspensions of fly samples were created by placing individual larva, pupa, and adult specimens into storage solutions of 70% aqueous ethanol. These suspensions were subsequently analyzed by dipping the closed end of a melting point capillary tube into the solution before presenting the coated surface in the open-air space between the ion source and mass spectrometer inlet of the DART-HRMS instrument. For each life stage within a species, a minimum of five independent samples were analyzed and averaged together, in replicates of five. Chemical standards, typically solids, were analyzed in a similar manner by dipping the closed end of a melting point capillary into the powder and holding it before the ion source for several seconds to acquire a spectrum.

Mass spectral data were acquired using a DART-SVP ion source (IonSense, Saugus, MA) coupled to a JEOL AccuTOF mass spectrometer (JEOL USA, Peabody, MA) in positive-ion mode. The gas heater temperature was set to 350 °C, and the DART ion source grid voltage was 50 V. The mass spectrometer settings were ring lens voltage, 5 V; orifice 1 voltage, 20 V; orifice 2 voltage, 5 V; and peak voltage, 400 V. The helium flow rate of the DART ion source was 2.0 L/min. Spectra were collected at a rate of 1 spectrum per s over the mass range m/z 40−800. The resolving power of the mass spectrometer was 6000 fwhm. Mass calibration was performed using polyethylene glycol (PEG 600). Data processing

DOI: 10.1021/acs.analchem.8b01704
Anal. Chem. 2018, 90, 9206−9217

Scheme 1. Steps in the Multivariate Statistical Analyses of Carrion Insect DART-HRMS Data



including background subtraction, calibration, and peak centroiding were performed using TSSPro3 software (Schrader Analytical Laboratories, Detroit, MI). Mass spectral analysis and elemental composition determination were performed using Mass Mountaineer (Mass-spec-software.com, RBC Software, Portsmouth, NH). Confirmation of the identities of amino acids detected in the DART mass spectra was accomplished by in-source collision-induced dissociation (CID) experiments performed on the ethanol fly sample suspensions, as previously described.[59]

Multivariate Statistical Analysis of DART-HRMS-Derived Chemical Fingerprints. The mass spectral data for all the samples were stored in text format and imported into MATLAB 9. 3. 0, R2017b Software (The MathWorks, Inc., Natick, MA, USA) for further processing. Masses in the range of m/z 43−392 were used. The steps are outlined in Scheme 1. Following DART-HRMS analysis (Step 1), the data in the form of two column matrices comprised of m/z and intensity values were binned with an in-house written MATLAB code so that common m/z values were aligned along the m/z dimension. The bin width was selected based on the resolution of the JEOL AccuTOF-MS instrument used (i.e., ±5 mmu of the calculated mass). Assessment of the optimal relative abundance threshold to use is an iterative process and is based on that which yields the best accuracy and interpretability of the results. Therefore, different values of this parameter representing 1%, 2%, and 3% were tested to assess the impact on discrimination. Through this testing, a 2% threshold was observed to provide the highest prediction accuracy. Therefore, in the binning process, peak intensity values were filtered by selecting an abundance threshold of 2%. The mass data matrices for each life stage were normalized using "autoscaling" in order to (1) set all variables to unit variance and (2) ensure that the correlation rather than the covariance was the basis of the analysis (i.e., Step 2). In Step 3, the minimum redundancy maximal relevance (mRMR)[60] feature selection method was applied to determine features that were the primary contributors to the ability to discriminate between groups. Following this data reduction step, the number of variables that remained were 50, 50, and 60 for the larva, pupa, and adult data, respectively. Then in Step 4, Sparse Discriminant Analysis (SDA),[61,62] a method that accomplishes simultaneous classification and feature selection, was applied to the new 2D matrix comprised of the number of samples versus the selected m/z values, to investigate the quality of discrimination and to identify the variables most important for discrimination, based on their SDA coefficients (i.e., the relative weights of variables in discrimination).

Implementation of the SDA method on larva (30 × 50), pupa (30 × 50), and adult (36 × 60) data matrices gave three coefficient vectors for each, indicated by the three SD axes for each life stage shown in Step 4 of Scheme 1. The variable coefficients which were acquired from the SDA yielded 30 × 25, 30 × 25, and 36 × 32 matrices for larvae, pupae, and adults, respectively. This treatment resulted in a data reduction of 50%. In these matrices, the first number represents the number of samples, and the second, the most heavily weighted m/z values, respectively. In Step 5, these matrices served as the data set used in a Kohonen (i.e., self-organizing) map model. A graphical interface was used to apply the Kohonen map algorithm, the details of which have been published.[63,64] Briefly in Step 5, the data matrices were partitioned into two parts: the validation set, comprised of randomly selected observations, and the training set which included the remaining observations. The model was trained and tested by performing 5-fold venetian blind cross-validation. This learning model was then tested using the validation set to predict the identity of the test samples. Unknown samples in the Kohonen map prediction step were evaluated based on the trained Kohonen weights, and their positioning within the neural network enabled their assignment to a consistent class. It should be noted that the discrimination and identification were also accomplished by

DOI: 10.1021/acs.analchem.8b01704
Anal. Chem. 2018, 90, 9206−9217

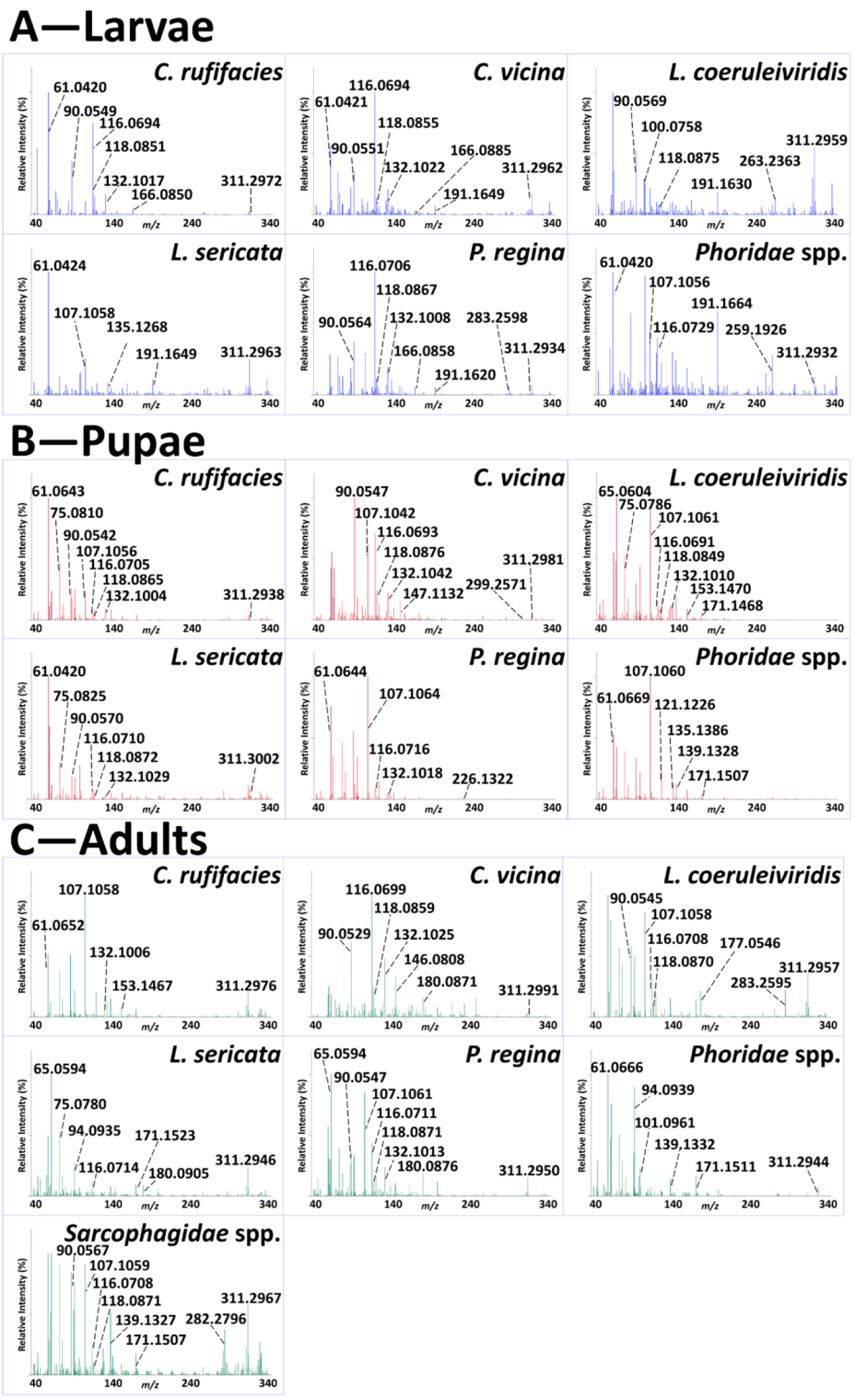


Figure 2. Results of DART-HRMS analyses of aqueous ethanol suspensions of the life stages of the indicated species. Panel A, larvae; panel B, pupae; panel C, adults. All analyses were performed in positive-ion mode at 350 °C. Each spectrum serves as a fingerprint for the indicated species and is an average of 5 analyses. The mass measurement data, including masses and relative peak abundances, are listed in Supporting Information Tables S1–S3.

PCA-DA and PLS-DA methods, and comparable results were obtained. However, we chose to use the Kohonen map model because it enabled us to also determine the weights of the variables that were important for defining a class (termed "marker variables").

RESULTS

Mass Spectrometric Analysis of Carrion Insect Larva, Pupa, and Adult Life Stages. *Larvae.* The ethanol suspensions of the larvae samples were analyzed by DART-HRMS in positive-ion mode at 350 °C. The results for the six fly larvae species analyzed (*C. rufifacies, C. vicina, L.*

DOI: 10.1021/acs.analchem.8b01704
Anal. Chem. 2018, 90, 9206–9217

coeruleiviridis, L. sericata, P. regina, and *Phoridae* spp.) are shown in Figure 2A and the relevant mass measurement data including masses and relative peak abundances are listed in the Supporting Information Table S1. Peaks at nominal m/z 47 and 93.09 were detected in the spectra of all six species and corresponded to ethanol and its dimer respectively, both in protonated form. As has been reported previously,[59] these peaks are derived from the aqueous ethanol solution in which the fly larvae were stored and were thus removed from all the spectra. Visual inspection of the mass spectra revealed a unique chemical fingerprint for each species. The spectra displayed between 31 and 155 peaks above a 2% relative abundance threshold; *C. rufifacies* had the least number of peaks and *Phoridae* spp. had the greatest.

Although high-resolution masses were obtained in these experiments, those that are highlighted henceforth are listed as nominal masses except in cases where m/z values with the same nominal mass differ by ≥6 mmu. In these cases, the observed high-resolution mass is listed. Despite the fact that each species displayed a unique chemical fingerprint, there were some peaks that were common to all species including m/z 61.04, 62.06, 72.08, 90.06, 101, 107.10, 116, and 311. Other peaks were observed in the majority of the spectra, such as m/z 61.07, 100.08, 191.16, and 333 which appeared in all the species except for *C. rufifacies*; m/z 60.08, 76.04, 86.09, and 118 that were observed in the spectra of all of the species but *L. sericata*; m/z 75, 139.13, and 255 which were detected in all of the species spectra except *P. regina*; and m/z 70 that was observed in all of the species except for both members of the *Lucilia* genus, i.e., *L. coeruleiviridis* and *L. sericata*.

Observed in all of the spectra were peaks at m/z 116 and 90.06, which corresponded to protonated proline and alanine, respectively. The peak at m/z 118 that was detected in the spectra of all of the species except for *L. sericata* was identified as valine $[H^+]$. Peaks at m/z 76.04 and 132 were detected in the mass spectra of *C. rufifacies, C. vicina, L. coeruleiviridis, P. regina*, and *Phoridae* spp. and corresponded to protonated glycine and isoleucine/leucine, respectively. Protonated phenylalanine (m/z 166) and protonated threonine (m/z 120) were present in the ethanol larvae suspensions of *C. rufifacies, C. vicina*, and *P. regina*. The peak at m/z 147.11 corresponding to lysine $[H^+]$ was detected in *C. vicina* and *P. regina*. Several peaks were only observed in one species. These included m/z 106 (corresponding to protonated serine) and m/z 150 (corresponding to protonated methionine), which was only observed in the *P. regina* larvae suspensions; m/z 147.07 corresponding to protonated glutamine, which was only detected in *Phoridae* spp.; and m/z 148 (protonated glutamic acid), which was only observed in *C. vicina*.

Pupae. The results of the DART-HRMS analyses of the aqueous ethanol suspensions of the six species of pupae samples (*C. rufifacies, C. vicina, L. coeruleiviridis, L. sericata, P. regina, and Phoridae* spp.) are shown in Figure 2B, and the relevant mass measurement data are listed in the Supporting Information Table S2. As was the case for the larvae ethanol suspensions, the spectra of the pupae samples displayed unique chemical fingerprints. Depending upon the species, between 25 and 61 peaks above a 2% relative abundance threshold were observed, with *Phoridae* spp. containing the least and *C. vicina* the most number of peaks.

Similar to what was observed for the larvae samples, there was a subset of peaks that were common to all of the pupae ethanol suspensions, such as m/z 48, 61.06, 64.08, 75.08, 79, 93.22, 93.27, 94.09, and 139. A number of peaks were observed in a subset of the species analyzed, including m/z 43 that appeared in the spectra of all the species except for *C. vicina*; m/z 61.04 which was detected in the spectra of all the species except for *L. coeruleiviridis*; peaks at m/z 89 and 107 that appeared in the spectra of all species except for *L. sericata*; m/z 65, 100.08, and 171 which were observed in all of the species except for *P. regina*; and m/z 90.06, 116, 118, and 132 which were detected in all of the species except for *Phoridae* spp.

As was the case for the larvae samples, several of the detected molecules were confirmed to be protonated amino acids. These included m/z 90.05, 116, 118, and 132 that were observed in *C. rufifacies, C. vicina, L. coeruleiviridis, L. sericata*, and *P. regina* and which corresponded to alanine, proline, valine, and isoleucine/leucine, respectively. A mass at m/z 76.04 was observed in *C. rufifacies, C. vicina*, and *P. regina*, which corresponded to glycine. Threonine at m/z 120 was observed of *C. vicina* and *L. coeruleiviridis* samples, while the peak at m/z 147.11 that appeared in the spectrum of *C. vicina* corresponded to lysine. There were no amino acids detected above a 2% relative threshold in the aqueous ethanol suspension of *Phoridae* spp. pupae.

Adults. The results of positive-ion mode DART-HRMS analysis at 350 °C of the seven adult species analyzed (*C. rufifacies, C. vicina, L. coeruleiviridis, L. sericata, P. regina, Phoridae* spp., and *Sarcophagidae* spp.) are illustrated in Figure 2C, and the related mass measurement data are listed in the Supporting Information Table S3. As was observed for the larval and pupal suspensions, the mass spectra displayed unique chemical fingerprints for each species. The number of peaks above a 2% relative abundance threshold ranged from 44 to 181. *C. rufifacies* had the least number and *Sarcophagidae* spp. had the greatest.

The results were similar to those observed in the larvae and pupae ethanol suspensions in that some of the m/z values were detected in all species while others only appeared in a subset. Peaks at m/z 61.06, 62.06, 64.08, 72.08, 75, 79, 94.09, 116, 139, and 311 were observed in all samples. Peaks at m/z 43, 95, and 171 were observed in all of the spectra except for *C. vicina*; and m/z 70, 90.06, 118, 132, and 309 were detected in the spectra of all the species except for *Phoridae* spp. Masses that were confirmed to be protonated amino acids were as follows: the peak detected in all seven species at nominal m/z 116 corresponded to proline; m/z 90.06, 118, and 132 observed in ethanol suspensions of *C. rufifacies, C. vicina, L. coeruleiviridis, L. sericata, P. regina*, and *Sarcophagidae* spp. corresponded to alanine, valine, and isoleucine/leucine, respectively; m/z 76.04 and 106 detected in *C. vicina, P. regina*, and *Sarcophagidae* spp. samples corresponded to glycine and serine, respectively; m/z 120, 150, and 166 observed in *C. vicina* and *P. regina* corresponded to threonine, methionine, and phenylalanine, respectively; m/z 147.08 and 156.08 in *C. vicina* corresponded to glutamine and histidine, respectively; and m/z 148 in *Sarcophagidae* spp. corresponded to glutamic acid. The profile of amino acids identified in each life stage is presented in Table S4.

As was observed in previous studies examining blow fly eggs,[59] some interesting trends were seen in the distribution of the amino acids detected in the aqueous ethanol suspensions of the larva, pupa, and adult samples analyzed. The amino acids arginine, asparagine, aspartic acid, cysteine, tryptophan, and tyrosine were not observed in the ethanol solutions of any life stage of any species. Alanine, glycine, isoleucine/leucine,

DOI: 10.1021/acs.analchem.8b01704
Anal. Chem. 2018, 90, 9206−9217

Table 1. Input and Output Properties of the Kohonen Artificial Neural Networks

Data set properties			Kohonen map properties[a]				Classification results		
Sample type	Training set	Validation set	Network size	Epoch	Learning rate	Class scaling factor	Model error	CV[b] error	Accuracy %
Larvae	25 × 25	5 × 25	5	200	0.03−0.005	0.9	0	0	100
Pupae	25 × 25	5 × 25	5	300		1	0	0.04	96
Adults	30 × 32	6 × 32	6	350		0.9	0	0.06	93

[a]The training algorithm was batch and the boundary condition was normal. [b]Cross-validation.

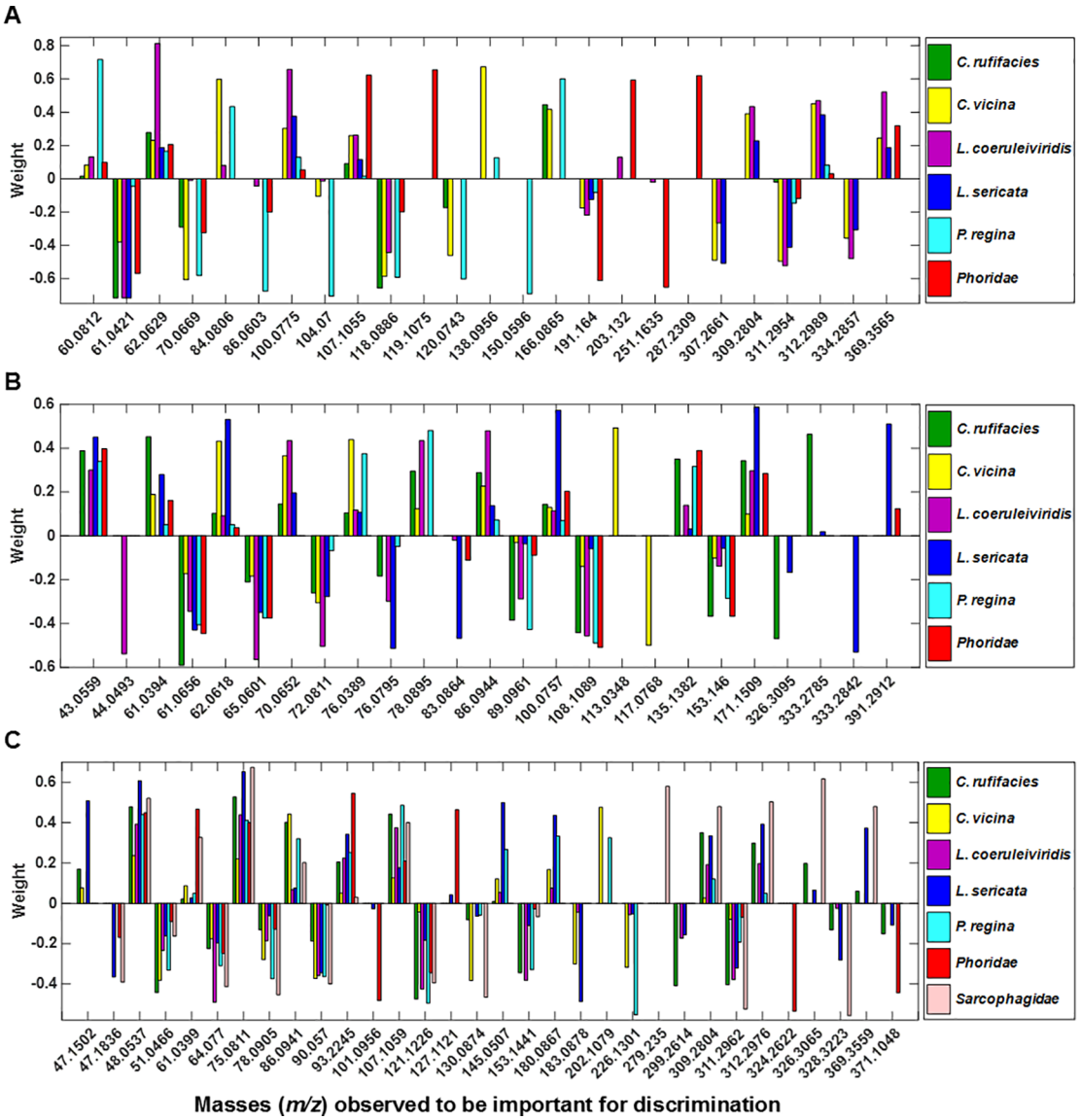


Figure 3. Color-coded profiles of Kohonen weights of the masses associated with the neurons in the artificial neural networks (ANNs) shown in Figure 5. The *x*-axis of each panel shows the subset of individual *m*/*z* values determined from the ANNs to be the most important for discriminating between the species. The presence of a bar indicates that the corresponding mass is observed in the indicated species, and the bar height shows how heavily weighted the corresponding *m*/*z* value is in discriminating between the represented species and the others. The weights for half of the indicated masses are made negative for ease of visualization. Therefore, the negative weights should be interpreted as their absolute values. Panel A, larvae; panel B, pupae; and panel C, adults.

proline, and valine were the most common amino acids detected across all of the species and life stages. The pupae and adult aqueous ethanol suspensions of *Phoridae* spp. did not appear to contain many amino acids, as only proline was detected within the adult samples, and no amino acids were observed in the pupae.

Species Identification Was Accomplished Through Processing of the DART-HRMS Chemical Fingerprint Profiles Using the Kohonen Map Artificial Neural Network. The statistical analysis processing of the DART-HRMS chemical profiles is outlined in Scheme 1. Following binning and normalization, the dimensionality of the data was reduced through the application of a series of successive steps,

DOI: 10.1021/acs.analchem.8b01704
Anal. Chem. 2018, 90, 9206−9217

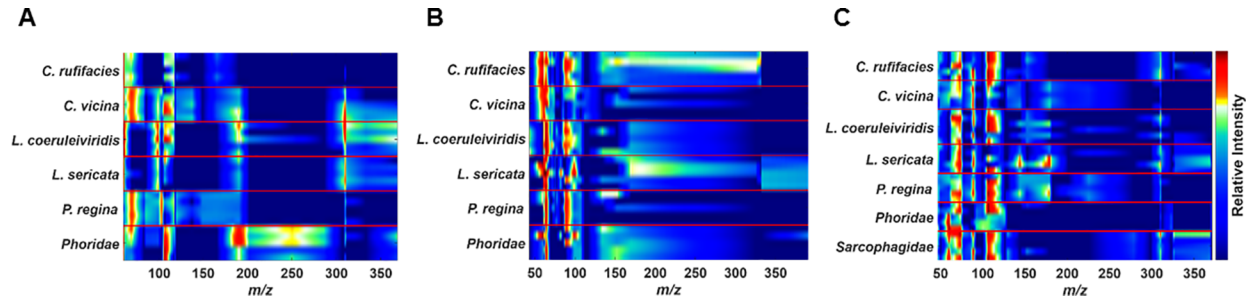
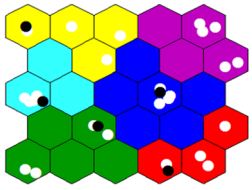


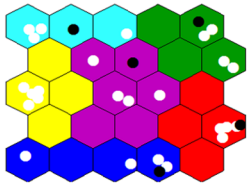
Figure 4. Heat map renderings of the subset of weighted m/z values revealed by the Kohonen artificial neural networks to be the most important for species discrimination and identification within each life stage. Panel A, larva mass spectral data representing replicates of 6 species (25 m/z values); panel B, pupa mass spectral data representing replicates of 6 species (25 m/z values); and panel C, adult mass spectral data representing replicates of 7 species (32 m/z values).

A—Larvae



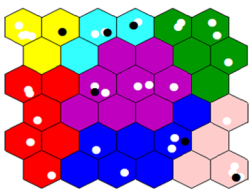
Classifier Performance					Validation Result****	
Species	Sensitivity*	Specificity**	Precision	AUC***	Species	Predicted class
	1	1	1	1		True
	1	1	1	1		True
	1	1	1	1		True
	1	1	1	1		True
	1	1	1	1		True
	1	1	1	1		

B—Pupae



Classifier Performance					Validation Result****	
Species	Sensitivity*	Specificity**	Precision	AUC***	Species	Predicted class
	1	1	1	1		True
	1	1	1	1		True
	1	1	1	1		True
	1	0.95	0.80	0.98		True
	1	1	1	1		True
	0.75	1	1	0.88		

C—Adults



Classifier Performance					Validation Result****	
Species	Sensitivity*	Specificity**	Precision	AUC***	Species	Predicted class
	0.80	1	1	0.90		True
	1	1	1	1		True
	0.80	0.96	0.80	0.86		True
	1	1	1	1		True
	1	0.96	0.75	0.98		True
	1	1	1	1		True
	1	1	1	1		

C. rufifacies *C. vicina* *L. coeruleiviridis* *L. sericata* *P. regina* *Phoridae* *Sarcophagidae*

*Sensitivity also refers to the true positive rate; **1 minus the specificity is equal to the false positive rate; ***AUC refers to area under the "receiver operator characteristic" (ROC) curve; ****Refers to external validation.

Figure 5. Results of multivariate statistical analyses of life stage data matrices for the larvae (panel A), pupae (panel B), and adults (panel C). The merits of the performance of the Kohonen models (i.e., sensitivity, specificity, precision, and area under ROC curve (AUC)), are presented in the tables. The panels show the top view of the Kohonen maps of the datasets, as well as where the training samples fell within the neural network (indicated by white circles), and where the external validation samples fell (indicated by black circles). The validation samples (5, 5, and 6 observations for the larvae, pupae and adults respectively) were randomly selected observations, and the training set comprised the remaining observations. For the larvae and pupae, there was no duplication of species in the randomly selected validation samples, while for the adults, two samples of the same species (i.e. *P. regina*) were included in the randomly selected samples. The prediction reults for the external validation samples are shown in the tables and indicate that all predictions were 100% accurate.

starting with minimum redundancy maximal relevance (mRMR) feature selection. This resulted in 30 × 50, 30 × 50, and 36 × 60 matrices representing the larva, pupa, and adult species, respectively. These matrices, which reflected the m/z values that contributed most to the variance in the data, were then subjected to Sparse Discriminant Analysis (SDA). This step resulted in the further selection of the variables that contributed most to discrimination, and translated each to a new dimensional space in an SDA plot. The coefficients of the variables that were most important for discrimination were then used to construct matrices whose dimensionality was further reduced. The matrices with the reduced number of variables were then partitioned into validation and training sets, with the training sets being comprised of 25 × 25, 25 × 25, and 30 × 32 matrices representing the larvae, pupae, and adults, respectively. The training set data served as the input for a Kohonen self-organizing map which was used as a discrimination model for blow fly species determination, and to reveal the most heavily weighted variables that contributed to the ability to distinguish between each species within a given life stage. The input and output properties of the resulting artificial neural network appear in Table 1. The matrix

DOI: 10.1021/acs.analchem.8b01704
Anal. Chem. 2018, 90, 9206−9217

Scheme 2. Confusion Matrices for the Pupa and Adult Models[a]

A—Pupae Confusion Matrix

		Pupae					
		Predicted					
	Class	*C. rufifacies*	*C. vicina*	*L. coeruleiviridis*	*L. sericata*	*P. regina*	*Phoridae*
True	*C. rufifacies*	4	0	0	0	0	0
	C. vicina	0	5	0	0	0	0
	L. coeruleiviridis	0	0	4	0	0	0
	L. sericata	0	0	0	4	0	0
	P. regina	0	0	0	0	4	0
	Phoridae	0	0	0	1	0	3

B—Adults Confusion Matrix

		Adults						
		Predicted						
	Class	*C. rufifacies*	*C. vicina*	*L. coeruleiviridis*	*L. sericata*	*P. regina*	*Phoridae*	*Sarcophagidae*
True	*C. rufifacies*	4	0	1	0	0	0	0
	C. vicina	0	4	0	0	0	0	0
	L. coeruleiviridis	0	0	4	0	1	0	0
	L. sericata	0	0	0	4	0	0	0
	P. regina	0	0	0	0	3	0	0
	Phoridae	0	0	0	0	0	5	0
	Sarcophagidae	0	0	0	0	0	0	4

Gray shading highlights a misclassification.

[a]A confusion matrix is not presented for the larva data because there were no misclassifications in the internal validation for that dataset.

dimensions of the training and validation sets for each life stage are shown and the network size was 5, 5, and 6 for the larvae, pupae, and adults, respectively. For training, initial estimates of Kohonen weights were selected on the basis of PCA eigenvectors. The epochs corresponding to each life stage (i.e., the number of times each sample was introduced into the network) were 200, 300, and 350, respectively, for larvae, pupae, and adults. A batch training algorithm and normal boundary conditions were used. The learning rate at the start of the training was 0.03 and at the end was 0.005. The 5-fold venetian blind cross-validation misclassification errors were 0, 0.04, and 0.06 for the larvae, pupae, and adults, respectively. Cross-validation showed an accuracy of 100, 96, and 93 for the larva, pupa, and adult species, respectively. The false positive and false negative rates for the larvae, pupae, and adults, respectively were (0, 0), (0.0476, 0.25), and (0.077, 0.4). The profile of trained Kohonen weights associated with each of the discrimination markers (i.e., m/z values) used is shown in Figure 3, panels A, B, and C for the larva, pupa, and adult species, respectively. In the bar charts featured in each panel, the colors green, yellow, magenta, blue, turquoise, and red correspond to *C. rufifacies*, *C. vicina*, *L. coeruleiviridis*, *L. sericata*, *P. regina*, and *Phoridae* spp., respectively. *Sarcophagidae* spp., an additional species represented in panel C (adults), is denoted in pink. The x-axis of each panel shows the m/z values determined from the self-organizing map to be the most important for discriminating between the species. The presence of a bar indicates that the corresponding mass is observed in the indicated species, and the bar height shows how heavily weighted the corresponding m/z value is in discriminating the species from the others. The weights for half of the indicated masses are made negative for ease of visualization. Therefore, the negative weights should be interpreted as their absolute values. Heat map renderings of the most heavily weighted masses that defined the species within each life stage are shown in Figure 4. It provides an illustration of the high similarity between replicates for the subset of masses that were important in species identification and discrimination for the species represented in each life stage. For example, for the larvae in panel A, the color gradations for the samples representing *C. rufifacies* are highly consistent. The same is true for *L. sericata* and *P. regina*. However, more variability was observed for *C. vicina* as illustrated (for example) by the variation in color intensity for masses above m/z 300.

The rendering of the data in the manner shown in Figure 3 revealed a number of trends. First, for each life stage, there was a subset of heavily weighted masses that were unique to a particular species. For example, for the larvae (panel A), nominal m/z 119 and 287 were unique to *Phoridae* spp. and m/z 150 was unique to *P. regina*. For the pupae (panel B), m/z 44 was unique to *L. coeruleiviridis*; m/z 113 and 117 were unique to *C. vicina*; and m/z 333 was unique to *L. sericata*. For the adults (panel C), m/z 279 was unique to *Sarcophagidae* spp.; and m/z 324 was unique to *Phoridae* spp. Within each life stage was a subset of masses that were weighted to differing extents for all species. These included m/z 61.04, 62.06, 107 and 311 for the larvae; m/z 61.06, 62.06, 65, 89, 100, 108, and 153 for pupae; and m/z 48, 51, 64, 75, 78, 90, 93, 107, 121 and 311 for adults. For each life stage, there were several m/z values that were shared by only a subset of species. Examples included m/z 84 that was shared by *C. vicina*, *L. coeruleiviridis* and *P. regina* larvae; m/z 326 that was shared by *C. rufifacies* and *L. sericata* pupae; and m/z 202 that was shared by *C. vicina* and *P. regina* adults.

DOI: 10.1021/acs.analchem.8b01704
Anal. Chem. 2018, 90, 9206−9217

The 2-D Kohonen map output layers for the trained larva, pupa, and adult models (which also correspond to the aforementioned Kohonen weights shown in Figure 3) are illustrated in Figure 5A–C. The merits of the performance of the models (i.e., sensitivity, specificity, precision, and area under the ROC "receiver operator characteristic" curve (AUC)), and the prediction results for the validation samples are presented in the tables. By convention, the maps appear as honeycombs comprised of individual neurons each in the shape of a hexagon, with like-colored neurons indicative of a specific class. The white circles within the Kohonen maps illustrate the position of the trained samples within the neurons, while the black circles show where within the neural network the external validation samples fell. The results of the cross validation predictions for the larvae were 100% accurate for all species, while a few misclassifications appeared for some species in the pupa and adult groups. For example, for the pupae, the sensitivity for *Phoridae* was 0.75, and the specificity and precision for *L. sericata* were 0.95 and 0.80, respectively. These values of less <1 reflect misclassifications, the details of which are shown in the confusion matrix in panel A of Scheme 2. For example, for the pupae, one *Phoridae* sample was misclassified as *L. sericata*. For the adults, the classifier performance presented in Figure 5 shows a sensitivity of 0.8 for both *C. rufifacies* and *L. coeruleiviridis*, a specificity of 0.96 for both *L. coeruleiviridis* and *P. regina*, and precision of 0.8 and 0.75 for *L. coeruleiviridis* and *P. regina*, respectively. These classifier performance parameters of <1 again reflect misclassifications, the details of which are presented in the confusion matrix in Scheme 2, panel B. It shows that a *C. rufifacies* sample was misclassified as *L. coeruleiviridis*, and a *L. coeruleiviridis* sample was misclassified as *P. regina*. However, despite the misclassifications observed in the cross validation, the results for the external validation (indicated in black circles in the neural network shown in Figure 5) were 100% accurate in all cases. Five samples from each of the species representing the larvae and pupae, and six from the adult species, are shown in the tables in Figure 5A–C (see "Validation Result"). In all cases, the validation samples were accurately predicted.

DISCUSSION

We demonstrate here that ethanol suspensions of samples of a given life stage for species within the *Calliphoridae*, *Phoridae*, and *Sarcophagidae* families exhibit similar DART-HRMS spectra, but that these differ between species for the same life stage, even for closely related species. For example, *L. sericata* and *L. coeruleiviridis* larvae are visually similar and difficult to distinguish. Samples of *L. coeruleiviridis* larvae hatched from eggs laid by several flies yield spectra that are highly similar, and this was also true of samples of *L. sericata*. However, despite both being members of the same genus, their DART-HRMS spectra were different from one another. Thus, for the species represented in this study, the observed chemical profiles showed consistent intraspecies similarities and interspecies differences. Inherent in the data were qualities that enabled accurate attribution of species identity to the mass spectra of sample unknowns. This was revealed through a supervised learning approach featuring a Kohonen model artificial neural network for each life stage. The results showed that these models were successful in predicting the identity of unknown species on the basis of selected m/z values using multivariate statistics (Table 1 and Figures 3 and 5). The profiles of the weighted Kohonen variables are indicated in bar plots in Figure 3 and reveal m/z values that were significant for the training of each class and for discrimination between classes. Previous publications (Brereton et al. and references cited therein)[65] have discussed the advantages of Kohonen mapping in modeling and visualization when compared to Principal Component Analysis Discriminant Analysis (PCA-DA) and Partial Least Squares Discriminant Analysis (PLS-DA). We confirmed this in our study by implementing both PCA-DA and PLS-DA methods, and we obtained results comparable to those observed in Kohonen mapping (data not shown). However, one of the most important uses of supervised self-organizing maps (SOMs) involves determining what variables are important for the purpose of defining a class or group of samples, and this gave the Kohonen approach an edge over PCA-DA and PLS-DA.

Examination of the SOMs in panels B and C of Figure 5 shows that the samples which in cross-validation steps were classified incorrectly and which affected the sensitivity and specificity results, always involved species that were neighbors in the SOMs and those that fell near the boundaries separating the species in the SOMs. For example, for the pupae, one *Phoridae* sample was classified as *L. sericata* (Scheme 2, panel A), and in the SOM for the pupae (Figure 5, panel B), it is illustrated that *Phoridae* and *L. sericata* are neighbors. Similarly, the confusion matrix for the adults (Scheme 2, panel B) shows that a *C. rufifacies* sample was classified as *L. coeruleiviridis*, and a *L. coeruleiviridis* sample was classified as *P. regina*. The species represented in each of these pairs are neighbors in the SOM shown in Figure 5; *C. rufifacies* (green) buttresses *L. coeruleiviridis* (magenta) and *L. coeruleiviridis* abuts *P. regina* (turquoise). Nevertheless, despite the deviation from unity of the specificity and sensitivity for some classes (e.g., sensitivity of 0.75 for *Phoridae* pupae or 0.8 for *C. rufifacies* adults), the area under the ROC curve (AUC) is close to 0.9 in all cases, indicating that the model is highly discriminating, and it shows the suitability of the models for identification of species. The AUC is >0.9 for species discrimination in all cases except for *Phoridae* pupae (AUC of ∼0.88) and *L. coeruleiviridis* (AUC of ∼0.86) in the adult model. Usually, AUC values of 0.5–0.7 are indicative of a low accuracy model, values of 0.7–0.9 reflect models that have useful applications, and values of >0.9 show high accuracy.[66] However, the risk for false identifications that may occur for samples near the edges of the boundaries between classes in the SOMs may be reduced through: (1) independent assessment of the similarity parameters (e.g., correlation or Euclidean distance, among others) for a given sample and class members; and (2) assessment of the presence of diagnostic m/z values (such as those presented in Figure 3).

A previous study showed that multivariate statistical analysis processing of the DART-HRMS spectra of necrophagous insect eggs could be used to determine species identity.[59] In that work, the ability to discriminate between species was a consequence of differences between their amino acid profiles. In this study, while intraspecies similarities and interspecies differences between amino acid profiles were also observed, the ability to differentiate and identify the species was based both on amino acid profiles and on masses whose identities are currently unknown. The structures of the unknowns are being actively investigated in our laboratory. We also observed that: (1) for a given species, there were no unique masses that were present throughout all the life stages and which distinguished it from other species; and (2) the earlier the life stage, the less complex were the spectra.

DOI: 10.1021/acs.analchem.8b01704
Anal. Chem. 2018, 90, 9206–9217

The species make up of blow fly populations is defined by area. For example, although there is some overlap, the range of blow fly species common to the Southwestern U.S. is different from that found in the Northeastern U.S. The species represented in this study, aside from *C. rufifacies*, include those that are common to the Manhattan area of New York City where they were collected with baited traps. Therefore, the artificial neural networks for larva, pupa, and adult blow fly species that were generated in this study could find immediate use in areas with a similar distribution of species. Importantly, the models can be continuously expanded to include not only more data representative of the species studied here, but also additional species found in other areas. This would provide a much needed comprehensive database against which the ethanol suspensions of DART-HRMS-derived chemical fingerprints can be screened for rapid species identification using life stages that cannot currently be used in this manner. Furthermore, the approach is rapid and utilizes samples in the form in which they are generated in the field (i.e., as suspensions in aqueous ethanol). An added advantage, particularly in a forensics context, is that the insect evidence itself is not destroyed, since it remains preserved in solution.

Our mass spectral analyses revealed several interesting trends including: (1) the ethanol suspension of each species and life stage exhibited a unique chemical fingerprint; (2) the chemical fingerprints were consistent between replicates of the same life stage and species, and differed considerably between species; (3) arginine, asparagine, aspartic acid, cysteine, tryptophan, and tyrosine were not detected in the spectra of any species or life stage. However, chemical standards of these amino acids were easily detected by DART-HRMS methods. This implies that if they were present, the levels of these amino acids within the aqueous ethanol suspensions were too low to be detected. A similar trend was also observed previously in the analysis of necrophagous blowfly eggs, in that the amino acids arginine, asparagine, and cysteine were not detected.[59] The significance of these observations as they pertain to insect biology is not currently known.

CONCLUSIONS

DART-HRMS-derived chemical fingerprint profiles representing the life stages of several species of carrion flies exhibit highly consistent intraspecies similarities and interspecies differences. When these data are used as the input for life-stage specific artificial neural networks based on the Kohonen self-organizing map model, a subset of masses which are most heavily weighted in enabling species predictions to be made were revealed. The resulting neural networks could be used for species-level identification of larvae, pupae, and adult samples of unknowns with predictive accuracies of 100%, 96%, and 93%, respectively, determined by cross-validation. The external validation accuracy was 100% in all evaluated cases. DART-HRMS analyses were performed on the aqueous ethanol suspensions of the samples, which is the form in which such samples are typically generated in the field, and no additional sample preparation is required. The artificial neural networks developed here provide the opportunity to rapidly identify several species within the Diptera order including those representing *Calliphoridae* (*C. rufifacies*, *C. vicina*, *L. coeruleiviridis*, *L. sericata*, and *P. regina*), *Phoridae*, and *Sarcophagidae*. The model can be expanded to include an increasing number of species and ultimately be used as a database against which DART-HRMS data can be screened to rapidly identify carrion insect species.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.8b01704.

Tables of mass measurement data of fly spectra and amino acid profiles of the various life stages (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: rmusah@albany.edu.

ORCID

Rabi A. Musah: 0000-0002-3135-4130

Author Contributions

R.A.M. and J.Y.R. conceived of the project. J.E.G. and J.Y.R. performed experiments. S.B. performed statistical analysis. J.Y.R. collected eggs in the field, reared and maintained fly colonies, and conducted fly species identification. R.A.M. wrote the manuscript. J.E.G. and S.B. contributed to writing the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Joey Fragale and Veena Mehta for insect care. The financial support of the Research Foundation of SUNY, the John Jay College for the Advancement of Research Faculty Scholarship Program and the United States National Institute of Justice is gratefully acknowledged. The project was supported in part by Award Number 2015-DN-BX-K057 awarded by the United States National Institute of Justice. The opinions, findings, and conclusions or recommendations expressed are those of the authors and do not necessarily reflect those of the Department of Justice.

REFERENCES

(1) Abela, G. *Br. J. Community Nurs.* **2017**, *22*, S14−S19.

(2) Byrd, J. H.; Castner, J. L. *Forensic Entomology: The Utility of Arthropods in Legal Investigations*; CRC Press: Boca Raton, FL, 2002; p 681.

(3) Greenberg, J. *A Natural History of the Chicago Region*, 2nd ed.; University of Chicago Press: Chicago, IL, 2002.

(4) Vasconcelos, S. D.; Soares, T. F.; Costa, D. L. *Int. J. Legal Med.* **2014**, *128*, 229−233.

(5) Reibe, S.; Madea, B. *Forensic Sci. Int.* **2010**, *195*, 52−57.

(6) Fratczak, K.; Matuszewski, S. *Forensic Sci. Int.* **2014**, *241*, 20−26.

(7) Pechal, J. L.; Moore, H.; Drijfhout, F.; Benbow, M. E. *Forensic Sci. Int.* **2014**, *245*, 65−71.

(8) Anderson, G. S. *J. Forensic Sci.* **2000**, *45*, 824−832.

(9) Tabor, K. L.; Brewster, C. C.; Fell, R. D. *J. Med. Entomol.* **2004**, *41*, 785−795.

(10) Tabor, K. L.; Fell, R. D.; Brewster, C. C. *Forensic Sci. Int.* **2005**, *150*, 73−80.

(11) Saigusa, K.; Matsumasa, M.; Yashima, Y.; Takamiya, M.; Aoki, Y. *Leg. Med.* **2009**, *11*, S344−S347.

(12) Sukontason, K.; Sribanditmongkol, P.; Ngoen-klan, R.; Klong-klaew, T.; Moophayak, K.; Sukontason, K. L. *Parasitol. Res.* **2010**, *106*, 641−646.

(13) Greenberg, B.; Szyska, M. L. *Ann. Entomol. Soc. Am.* **1984**, *77*, 488−517.

DOI: 10.1021/acs.analchem.8b01704
Anal. Chem. 2018, 90, 9206−9217

(14) Liu, D.; Greenberg, B. *Ann. Entomol. Soc. Am.* **1989**, *82*, 80−93.
(15) Greenberg, B.; Singh, D. *J. Med. Entomol.* **1995**, *32*, 21−26.
(16) Bourel, B.; Callet, B.; Hedouin, V.; Gosset, D. *Forensic Sci. Int.* **2003**, *135*, 27−34.
(17) Bala, M.; Sharma, A. *Egyptian J. of Forensic Sci.* **2016**, *6*, 203−208.
(18) Amendt, J.; Campobasso, C. P.; Gaudry, E.; Reiter, C.; LeBlanc, H. N.; Hall, M. J. *Int. J. Legal Med.* **2007**, *121*, 90−104.
(19) Grzywacz, A.; Szpila, K.; Pape, T. *Microsc. Res. Tech.* **2012**, *75*, 955−967.
(20) Moore, H. E.; Butcher, J. B.; Adam, C. D.; Day, C. R.; Drijfhout, F. P. *Forensic Sci. Int.* **2016**, *268*, 81−91.
(21) Day, D. M.; Wallman, J. F. *Forensic Sci. Int.* **2006**, *159*, 158−167.
(22) Wells, J.; LaMotte, L. *J. Forensic Sci.* **1995**, *40*, 585−590.
(23) Sadler, D. W.; Fuke, C.; Court, F.; Pounder, D. J. *Forensic Sci. Int.* **1995**, *71*, 191−197.
(24) Carvalho, L. M. L.; Linhares, A. X.; Trigo, J. R. *Forensic Sci. Int.* **2001**, *120*, 140−144.
(25) de Carvalho, L. M. L.; Linhares, A. X.; Badan Palhares, F. A. *Forensic Sci. Int.* **2012**, *220*, 27−32.
(26) Magni, P. A.; Pacini, T.; Pazzi, M.; Vincenti, M.; Dadour, I. R. *Forensic Sci. Int.* **2014**, *241*, 96−101.
(27) Adams, Z. J.; Hall, M. J. *Forensic Sci. Int.* **2003**, *138*, 50−61.
(28) George, K. A.; Archer, M. S.; Green, L. M.; Conlan, X. A.; Toop, T. *Forensic Sci. Int.* **2009**, *193*, 21−25.
(29) Valder, S. M.; Hopkins, T. L. *Ann. Entomol. Soc. Am.* **1968**, *61*, 827−834.
(30) Day, D. M.; Wallman, J. F. *Forensic Sci. Int.* **2008**, *179*, 1−10.
(31) Sukontason, K.; Sukontason, K. L.; Piangjai, S.; Boonchu, N.; Kurahashi, H.; Hope, M.; Olson, J. K. *Micron* **2004**, *35*, 391−395.
(32) Sukontason, K.; Sukontason, K. L.; Piangjai, S. *Rev. Inst. Med. Trop. Sao Paulo* **2003**, *45*, 95−98.
(33) Whitworth, T. L.; Dawson, R. D.; Magalon, H.; Baudry, E. *Proc. R. Soc. London, Ser. B* **2007**, *274*, 1731−1739.
(34) Chen, M. S.; Wheeler, S.; Davis, H.; Whitworth, R. J.; Knutson, A.; Giles, K. L.; Royer, T. A.; Skinner, M. *J. Econ. Entomol.* **2014**, *107*, 1110−1117.
(35) Wallman, J. F.; Donnellan, S. C. *Forensic Sci. Int.* **2001**, *120*, 60−67.
(36) Wells, J. D.; Stevens, J. R. *Annu. Rev. Entomol.* **2008**, *53*, 103−120.
(37) Malgorn, Y.; Coquoz, R. *Forensic Sci. Int.* **1999**, *102*, 111−119.
(38) Rolo, E. A.; Oliveira, A. R.; Dourado, C. G.; Farinha, A.; Rebelo, M. T.; Dias, D. *Forensic Sci. Int.* **2013**, *228*, 160−164.
(39) Mazzanti, M.; Alessandrini, F.; Tagliabracci, A.; Wells, J. D.; Campobasso, C. P. *Forensic Sci. Int.* **2010**, *195*, 99−102.
(40) Wells, J. D.; Williams, D. W. *Int. J. Legal Med.* **2006**, *121*, 1−8.
(41) Ames, C.; Turner, B.; Daniel, B. *Int. Congr. Ser.* **2006**, *1288*, 795−797.
(42) Zhu, G. H.; Xu, X. H.; Yu, X. J.; Zhang, Y.; Wang, J. F. *Forensic Sci. Int.* **2007**, *169*, 1−5.
(43) Moore, H. E.; Adam, C. D.; Drijfhout, F. P. *J. Forensic Sci.* **2013**, *58*, 404−412.
(44) Musah, R. A.; Espinoza, E. O.; Cody, R. B.; Lesiak, A. D.; Christensen, E. D.; Moore, H. E.; Maleknia, S.; Drijfhout, F. P. *Sci. Rep.* **2015**, *5*, 11520.
(45) Lavine, B. K.; Vora, M. N. *J. Chromatogr., A* **2005**, *1096*, 69−75.
(46) Page, M.; Nelson, L. J.; Blomquist, G. J.; Seybold, S. J. *J. Chem. Ecol.* **1997**, *23*, 1053−1099.
(47) Ye, G.; Li, K.; Zhu, J.; Zhu, G.; Hu, C. *J. Med. Entomol.* **2007**, *44*, 450−456.
(48) Yew, J. Y.; Cody, R. B.; Kravitz, E. A. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 7135−7140.
(49) Kim, H. J.; Seo, Y. T.; Park, S.; Jeong, S. H.; Kim, M. K.; Jang, Y. P. *Metabolomics* **2015**, *11*, 64−70.
(50) Giffen, J. E.; Lesiak, A. D.; Dane, A. J.; Cody, R. B.; Musah, R. A. *Phytochem. Anal.* **2017**, *28*, 16−26.
(51) Lesiak, A. D.; Cody, R. B.; Dane, A. J.; Musah, R. A. *Anal. Chem.* **2015**, *87*, 8748−8757.
(52) Lesiak, A. D.; Cody, R. B.; Dane, A. J.; Musah, R. A. *Forensic Sci. Int.* **2014**, *242*, 210−218.
(53) 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.
(54) Cody, R. B.; Dane, A. J.; Dawson-Andoh, B.; Adedipe, E. O.; Nkansah, K. *J. Anal. Appl. Pyrolysis* **2012**, *95*, 134−137.
(55) Maric, M.; Harvey, L.; Tomcsak, M.; Solano, A.; Bridge, C. *Rapid Commun. Mass Spectrom.* **2017**, *31*, 1014−1022.
(56) Vaclavik, L.; Cajka, T.; Hrbek, V.; Hajslova, J. *Anal. Chim. Acta* **2009**, *645*, 56−63.
(57) 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.
(58) Lopez-Esclapez, R.; Garcia, M. D.; Arnaldos, M. I.; Presa, J. J.; Ubero-Pascal, N. *Micron* **2014**, *62*, 43−51.
(59) Giffen, J. E.; Rosati, J. Y.; Longo, C. M.; Musah, R. A. *Anal. Chem.* **2017**, *89*, 7719−7726.
(60) Peng, H.; Long, F.; Ding, C. *IEEE Trans. Pattern Anal. Machine Intell.* **2005**, *27*, 1226−1238.
(61) Clemmensen, L.; Hastie, T.; Witten, D.; Ersbøll, B. *Technometrics* **2011**, *53*, 406−413.
(62) Sjöstrand, K.; Clemmensen, L. H.; Larsen, R.; Einarsson, G.; Ersbøll, B. *J. Stat. Softw.* **2018**, *84*, 1−37.
(63) Ballabio, D.; Consonni, V.; Todeschini, R. *Chemom. Intell. Lab. Syst.* **2009**, *98*, 115−122.
(64) Ballabio, D.; Vasighi, M. *Chemom. Intell. Lab. Syst.* **2012**, *118*, 24−32.
(65) Brereton, R. G. *Chem. Cent. J.* **2012**, *6*, S1−S1.
(66) Swets, J. A. *Science* **1988**, *240*, 1285−1293.

NOTE ADDED AFTER ASAP PUBLICATION

This paper published ASAP on 7/9/2018. The image for Figure 5 was replaced and the revised version was reposted on 7/10/2018.

DOI: 10.1021/acs.analchem.8b01704
Anal. Chem. 2018, 90, 9206−9217