

# Rapid Geographical Indication of Peppercorn Seeds Using Corona Discharge Mass Spectrometry

**Preeyarad Charoensumran**

Chulalongkorn University

**Monrawat Rauytanapanit**

Chulalongkorn University

**Nontawat Sricharoen**

Chulalongkorn University

**Barry L. Smith**

University of Liverpool

**Kanet Wongravee**

Chulalongkorn University

**Simon Maher**

University of Liverpool

**Thanit Praneenarat** (✉ [Thanit.P@chula.ac.th](mailto:Thanit.P@chula.ac.th))

Chulalongkorn University

---

## Research Article

**Keywords:** ambient ionization, mass spectrometry, corona discharge, linear discriminant analysis, geographical origin, black pepper

**Posted Date:** May 10th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-490348/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Scientific Reports on August 9th, 2021. See the published version at <https://doi.org/10.1038/s41598-021-95462-0>.

# Abstract

With increasing demands for more rapid and practical analyses, various techniques of ambient ionization mass spectrometry have gained interest due to its speed of analysis and the abundance of information provided. Herein, an ambient ionization technique that utilizes corona discharge was applied, for the first time, to analyze whole seeds of black and white peppers. This setup requires no solvent application or gas flow, thus resulting in a very simple and rapid analysis that can be applied directly to the sample without any prior workup or preparation. Combined with robust data pre-processing and subsequent chemometric analyses, this analytical method was capable of indicating the geographical origin of each pepper source with up to 98% accuracies in all sub-studies. The simplicity and speed of this approach open up the exciting opportunity for onsite analysis without the need for a highly trained operator. Furthermore, this methodology can be applied to a variety of spices and herbs, whose geographical indication or similar intellectual properties are economically important, hence it is capable of creating tremendous impact in food and agricultural industries.

## 1. Introduction

Mass spectrometry in combination with hyphenated techniques is the gold standard for chemical identification where legal enforcement is required<sup>1</sup>. Recently ambient ionization (AI) MS has been used to provide rapid assessment of authenticity, such as for the adulteration of Whisky<sup>2</sup>, pesticide residues on strawberry surfaces<sup>3</sup>, and bisphenol A and its analogs in food packaging<sup>4</sup>, to mention a few. AI approaches seek to reduce or entirely negate sample preparation, allowing analysis to be performed on samples in their native environment<sup>5</sup>. Popular AI sources include the likes of paper spray<sup>6-10</sup>, direct analysis in real time<sup>11-13</sup> and desorption electrospray<sup>14,15</sup>. AI sources that utilize corona discharge (CD) have been extensively reported, such as ASAP<sup>16,17</sup> and DAPCI<sup>18,19</sup>. These approaches typically rely on a heated nebulizing/desolvating gas flow in one form or another to vaporize a liquid/solid sample prior to ionization and analysis. CD without any accompanying gas flow is routinely used with ion mobility instruments<sup>20,21</sup> for trace vapor detection and has been incorporated with various AI sources to improve ionization efficiency. For instance, Song *et al.* applied CD in combination with an inexpensive ultrasonic nebulizer to detect antibiotic drugs in milk<sup>22</sup>, Mullen *et al.* combined secondary electrospray with CD to improve detection of explosive compounds<sup>23</sup>, and Sekimoto *et al.* increased DART ionization efficiency by addition of a CD pin between the source and inlet<sup>24</sup>. A limited number of studies have utilized direct ambient CD ionization for MS studies<sup>25,26</sup>.

A key aim of this study was to develop a methodology for direct peppercorn seed analysis that is sufficiently robust and does not require any sample preparation nor requires the addition of any consumables, such as solvents or gas. We surmised that CD ionization can be applied to solid phase samples that naturally contain volatile compounds for direct analysis without any prior vaporization. This significantly reduces the complexity of the setup to be simply the application of a CD needle in proximity to the sample. As is shown in this study, with appropriate sample types, the information obtained from

this analytical setup can have great impact to their respective fields. In this regard, aromatic spices are a perfect sample for this exploration because of the possible impact to the food and agricultural industry. For example, black pepper (*Piper nigrum*), colloquially called “King of Spices”, has evolved to be one of the most widely used spices in the world. Originating in India, black pepper has become a mainstay in both eastern and western cuisines, with its global export value reaching 1.6 billion USD in 2019<sup>27</sup>. Thus, various aspects of quality control of this product are vital to the sustainability of its market<sup>28–31</sup>. A prominent example is the development of analytical methods<sup>32,33</sup> to detect adulteration by cheaper materials, *e.g.*, pepper husk or papaya seed. Perhaps another important aspect of the quality control is geographical indication, which has become an increasingly important leveraging tool in trade negotiations<sup>34,35</sup> and sales valuation of products. GI protection is worth, on average, double the value compared to non-certified products<sup>36</sup>. In the EU alone, the value of GI products is estimated to be worth over €75 billion<sup>36</sup>. Therefore, the incentive for mis-characterization of product origin is very appealing for illegitimate producers. The drive to enforce GI necessitates the development of chemical analysis methods that are rapid, non-biased and versatile<sup>37–39</sup>.

Herein, we demonstrate, for the first time, the application of a CD for the direct ionization of peppercorn seeds to obtain a set of mass spectrometry (MS) data that are indicative of their geographical origins. Combined with robust pre-processing and chemometric analyses (principal component analysis (PCA) and linear discriminant analysis (LDA)), this analytical method was able to distinguish the geographical origin of different sources of black peppers, along with some simpler studies including the type of peppers (black vs white), and the origin of white peppers. An iterative reformulation feature selection algorithm was used to ascertain the key  $m/z$ 's necessary to construct accurate classifier models. Our findings indicate that this rapid, versatile and non-targeted MS protocol has the potential to be used for further GI studies relating to other herbs, spices and foodstuffs in general.

## 2. Material And Methods

### 2.1 Sample Preparation

Black pepper seeds originating from Thailand (6 sources), China (1 source), India (1 source), and the United Kingdom (3 sources) were used in this study; all samples were purchased locally from supermarkets or ordered online from reputable suppliers. White peppers from Thailand (6 sources) and China (1 source) were also purchased in a similar manner. All samples were removed from their respective packaging for analysis. Otherwise, no sample pretreatment or preparation was carried out. All experiments were performed in accordance with relevant guidelines and regulations.

### 2.2 Acquisition of MS data from peppers

A Waters Xevo TQ MS was used in this study with the following instrument parameters: cone voltage at 20 V, source temperature at 100°C, and acquisition mass range of 50–500  $m/z$ . The experimental method is depicted in Fig. 1. Briefly, a peppercorn seed was affixed with a distance of approximately 3 mm from

the MS inlet. A stainless steel needle was placed 2–3 mm above the sample with an applied potential of 3 kV. A corona discharge (CD) develops at the apex of the needle due to the large potential and small radius of curvature which subsequently ionizes vapors from the pepper sample. Data was acquired for a duration of 2 minutes with 1 second per scan. At least 9 seeds from each source were tested, with a blank (no seed) being run in between different sample sources. The complete set of data can be found in the supplementary data.

## 2.3 Data pre-processing for subsequent chemometric analyses

A series of automated data pre-processing steps were performed in MATLAB (Mathworks). (1) Firstly, any  $m/z$  values over 300 were discarded due to the lack of any significant ion peaks. (2) Mass spectra were smoothed with the weighted linear regression function within the peak width (10 data points) in order to prevent peak fluctuations in the same  $m/z$  channel. Baseline correction, with zeroing of any point with negative values, was performed. (3) The intensity of each  $m/z$  data point was scaled logarithmically in order to give each data point equal significance. (4) The data was aligned using 4 reference peaks ( $m/z = 109.1, 137.15, 177.15$  and  $235.3$ ) which are consistently present in all samples and background scans to ensure any instrument drift during analysis is corrected. (5) The MS was down sampled to 0.1 amu grid reducing the  $m/z$  datapoints from 3172 to 2000. (6) Local maxima within 0.8-amu windows were confirmed reducing the variables from 2000 to 267. (7) Peak intensity data from real samples were subtracted by the intensity data of the blank, *i.e.*, the runs without any peppercorn seed but with the same setup. (8) Finally, white and black pepper datasets are combined into a single data table consisting of 267 ( $m/z$ 's) by 156 (sample intensities) datapoints. Further filtering of  $m/z$ 's closer than 0.5 amu reduces the dataset to a final 199x156 datapoints. An overview scheme of the data preparation with an example mass spectrum after preprocessing is illustrated in Fig. S1.

## 2.4 Visualizing method and classification method

The data sets showed a very high number of variables (267 variables). To visualize and determine the number of significant factors from such high numbers, further data reduction was performed. The data was centered over all samples prior to the data reduction process. In this study, Principal Component Analysis (PCA) was used to reduce the dimensionality in a multivariate data set to those data characteristics that contribute the most to the overall variance, with the first component representing the greatest variance of the data (principal component 1 – PC1), the second greatest variance being projected to the second coordinate (PC2), and so on. PCA scores can be used to visualize the clusters of samples, which share common influences. More detailed information on PCA is discussed elsewhere<sup>40,41</sup>. In this study, PCA was employed to visualize the underlying relationships of the data. The clusters of samples were either visualized using PCs with the maximum variances (PC1-PC3), or using the best discriminating PCs.

To obtain the classification performance, Linear Discriminant Analysis (LDA) was used for class prediction. This was done by creating a model boundary (classifier) between classes using linear

discriminant function in order to define the directions in which the classes are best separated. However, in the LDA algorithm, it is necessary to calculate the inverse of the variance-covariance matrix. Therefore, if the number of variables is larger than the number of samples, the variance-covariance matrix will be a singular matrix that cannot be inverted. Hence, in this study, the discriminant features of the PCA-LDA were used as inputs to the classification model. The analysis was done by merging the modeling runs of two algorithms based on PCA to reduce the dimensionality of data matrix into 3PCs and sequentially LDA to create the classifier for class predictions of samples. The predictive ability of the developed PCA-LDA model is evaluated by “leave-one-out” cross validation (LOOCV). The mean centering is performed on the corresponding training set samples as appropriate, while the test sets are centered according to parameters obtained from the training set. The combination of validation procedures of PCA-LDA and the LOOCV approach were performed by the following steps:

1. For the data matrix ( $X$ ), a mass spectrum of a sample ( $x_{\text{test}}$ ) was removed to be used as a test set, while the remaining mass spectra are formed as a training set ( $X_{\text{train}}$ ).
2. PCA was performed on the training set to obtain score ( $T_{\text{train}}$ ) and loading ( $P_{\text{train}}$ ). In this study, only the first 3 PCs were used to represent the overall variance of the data.
3. The score ( $t_{\text{test}}$ ) of a test sample ( $x_{\text{test}}$ ) was calculated by using the pseudo-inverse of loading matrix from the training set ( $P_{\text{train}}$ ) as  $t_{\text{test}} = x_{\text{test}} P_{\text{train}}^T (P_{\text{train}} P_{\text{train}}^T)^{-1}$ , where  $T$  is a transpose operator and  $^{-1}$  is an inverse operator
4. The LDA classifier boundary of each class was built from  $T_{\text{train}}$  to predict the class of a test sample using  $t_{\text{test}}$ .

The procedure is repeated until all samples have been assigned as a test sample. The contingency table can be then constructed to express the performance and stability of the developed classifier.

## 2.5 Methods for identifying sets of significant variables

### 2.5.1 Fisher weight

Fisher weight<sup>42</sup> is defined by  $f_i = [\sum_{g=1}^G I_g (\bar{x}_{jg} - \bar{x}_j)^2] / [S_{\text{pool},j}^2 \sum_{g=1}^G (I_g - 1)]$  when there are  $G$  classes, where  $\bar{x}_j$  is the mean of variable ( $m/z$ )  $j$  of all classes,  $I_g$  is the number of samples in class  $g$ ,  $S_{\text{pool},j}^2$  the pooled standard deviation. The calculation is based on the ratio of within-class variance to between-class variance. It shows a main advantage over  $t$ -statistic as it can be further use for the data with  $> 2$  classes. However, when it is simplified for a two-class model ( $G = 2$ ), the rank of the variables is the same as the  $t$ -statistic although there is no sign. The  $m/z$  values with the highest values of fisher weight (rank number 1) are considered as potential candidates to be used as geographical markers.

### 2.5.2 Iterative reformulation of training sets

This technique<sup>43</sup> was performed by randomly selecting 70 % of the entire dataset as a training set for several times (100 in this case). Each training may thus result in different significant  $m/z$  values being selected. If the same  $m/z$  value is chosen many times as a geographical marker for each split in the data, it indicates that such a data point is likely a robust and authentic marker. In each iteration, the optimized number of ranked data points on Fisher weight scores are recorded as the most significant variables. The procedure was repeated 100 times, thus obtaining 100 lists of significant  $m/z$  data points. Any  $m/z$  values which presented in all iterations were identified as impactful geographical markers.

## 3. Results And Discussion

### 3.1 Features of MS data obtained from direct ionization of black peppers

An example spectrum (Fig. 2) highlighted some interesting features from the MS data obtained. First, it is surprising, at first glance, that a peak corresponding to piperine ( $m/z$  286 for  $[M + H]^+$ ) or its derivatives cannot be seen at all. Piperine is responsible for the pungency property of black pepper, and makes up to 7% of its dried weight<sup>44</sup>. This disappearance is likely due to the higher melting point and boiling point, as well as the higher polarity, compared to volatile terpenes, which is another group of compounds commonly found in black pepper. In fact, an aqueous extract of black pepper was shown to provide a clear signal of piperine via PS-MS (Fig. S2). Since the main goal of this work is to discover a key set of data that can differentiate the origins of black peppers, the absence of a single, albeit major, compound is deemed to be inconsequential. Next, volatile terpenes were indeed clearly found in this experimental setup. For example, a peak corresponding to monoterpene ( $C_{10}H_{16}$ ,  $m/z$  137 for  $[M + H]^+$ ) can be seen. Apart from piperine, these are a group of terpenes that directly contribute to the flavor profiles of black peppers<sup>45-48</sup>. Examples include myrcene, sabinene, and terpinene. Also, another peak at  $m/z$  205 is visible, which is attributed to sesquiterpene ( $C_{15}H_{24}$ ), with caryophyllene as a prominent example.

Importantly, the amounts of these compounds in each source of black pepper, as reflected in the ion intensities, can vary. For instance, while the representative sample in Fig. 2 had  $m/z$  151, likely carvone, as the most intense peak, some other samples had other peaks as the major component (Fig. S3 and S4 for the complete set of MS spectra, which were plotted from the complete set of numerical data in the supplementary data). These differences, when treated with systematic statistical analysis, may be sufficient for the differentiation of black peppers from different origins.

### 3.2 Data analysis of obtained MS data

In this study, multivariate data analysis was performed on the mass spectra in order to discriminate peppers based on three categorizations: I) the types of peppers (black and white peppers including all origins), II) the origins of black peppers (Thailand, China, India and UK), and III) the origins of white peppers (Thailand and China). First, one-sample  $t$ -tests were used to analyze the raw data for their variabilities in the analysis. It was found that all  $m/z$  data points from seeds of the same source do not

show any statistical difference at the 5% significance level, *i.e.*, accepting the null hypothesis. Thus, this indicates that variability of analysis was generally low. Then, all  $m/z$  data points were used for PCA analyses on the types of peppers, the origins of black peppers, and the origins of white peppers using the first three PCs (Fig. 3) with the total variance > 70% in all cases. Overall, the PCA score projections in this manner exhibit poor separation between classes in most studies. In addition, the contingency table of the classification using PCA-LDA is shown in Table S1. Classification rates of 76.92%, 74.46% and 69.35% were observed in the discrimination of the types of pepper, the origins of black peppers and the origins of white peppers, respectively. This result indicates that PCA based on the largest principal components may not always be the most effective method. In some cases, especially in biological systems<sup>43,49</sup>, the PCs with high variance may instead correlate with the background noise of biological samples and the best discriminator might present at the latter PC with small variance. Hence, to explore the possibility of obtaining useful information in small-intensity data points, some analysis on other PCs was conducted.

As shown in Fig. 4 (A,C,E), it can be seen that the classification accuracy using the first three PCs is poor for all three studies. With the inclusion of latter PCs, it was revealed that 18, 14, and 11 PCs were optimal for the classification of the types of peppers, the origins of black peppers, and the origins of white peppers, respectively. In addition, the prediction strength of each PC was calculated to visualize the discrimination of the PCs in each study (shown in Fig. S5). The higher the prediction strength, the better the discrimination power. It can be seen that some latter PCs showed higher prediction strength. That is, PC 3-4-10, PC 3-4-7, and PC 1-7-8, are the sets of PCs with the highest prediction strength for the studies of the types of peppers, the origins of black peppers, and the origins of white peppers, respectively. The score plots with the PCs with the highest prediction strength are shown in Fig. 4 (B,D,F).

To provide some numerical data for comparison, the classification rates for each PCA-LDA model built from both the first three PCs and the optimal number of PCs are illustrated in Fig. 5 (Table S1 for contingency tables). The overall classification accuracies using the first three PCs are merely 76.92%, 44.68% and 72.58% for the studies of the type of peppers, the origins of black peppers, and the origins of white peppers, respectively. On the other hand, the classification rates for the cases that were built from the optimal number of PCs appear to be significantly higher, with the scores of 98.72 %, 98.94 % and 100 %. Although the most optimal numbers of PCs for the LDA classifications were > 10 PCs in all cases, the classification rates increased dramatically only in the range of PC1-5. These results suggested that the first five PCs are the most relevant, but the latter PCs are still meaningful in differentiating between the classes. This is in good agreement with the PC score plots in Fig. 4 using the latter PCs with high prediction strength.

### 3.3 Significant $m/z$ data points as markers

A conventional approach for variable selection is to perform a selection method (Fisher weight in this case) on the entire dataset, and determine the significance of an  $m/z$  value as a marker from its magnitude of test value. However, selecting a set of markers from an entire dataset may encounter an overfitting issue. An alternative approach is to identify potential  $m/z$  data points from several spitted

training sets. This is called as “iterative reformulation of training set models”, which was first introduced elsewhere<sup>43</sup>. In this study, we also employed this approach to confirm the relevance of the obtained  $m/z$  data points. That is, a randomized partial set (70 % of the entire dataset) was used for the discovery of relevant  $m/z$  data points, whose appearances were then counted. After 100 iterations, all  $m/z$  values were then evaluated for the number of times they appeared in the list of relevant  $m/z$  data points of each split training set. As different  $m/z$  values may be differentially selected in each iteration, those data points that are most frequently selected are likely the most significant markers.

Figure 6 illustrates the number of times (out of 100 iterations) that each  $m/z$  value was selected in the subset of the top 10%  $m/z$  data points with the highest magnitude of Fisher weight value. Any  $m/z$  values that presented 100 times were identified as impactful markers. The study on the types of peppers provides a set of 3  $m/z$  data points ( $m/z = 205.2, 206.2$  and  $295.2$ ) that appeared in all 100 iterations (Fig. 6A), while the study about the origins of white peppers gave only 6  $m/z$  data points ( $m/z = 163.1, 205.2, 206.2, 222.2, 224.2,$  and  $238.3$ ) that appeared 100 times (Fig. 6C). In the case of the study about the origins of black peppers, there were 7  $m/z$  values ( $121.1, 205.2, 206.2, 245.2, 273.3, 291.2, 292.1,$  and  $292.2$ ) that appeared in all 100 iterations (Fig. 6B), which indicated a more challenging case due to higher numbers of variables. It can be seen that in all cases,  $m/z 205.2$  appeared as a significant marker, which is attributed to sesquiterpenes, a major class of compounds found in peppercorn seeds. It should be noted, however, that focusing on individual data points may be misleading due to the possibility for strategic adulteration/modification. This is actually reflected in the results, where different studies required different numbers of variables for effective cluster separations. Therefore, the non-targeted approach<sup>50</sup>, as practiced in this study, is deemed to be more flexible and versatile for broader applications. This is because it is not subject to strategic adulteration from a selected small set of known markers.

## 4. Conclusion

In this study, a simple and rapid MS-based method to reveal chemical profiles that are capable of distinguishing geographical origins was demonstrated. In particular, corona discharge was used to ionize volatile compounds from peppercorns for subsequent MS analysis. With its relatively simple setup, the method allowed for rapid data collection, which can then be further processed by chemometrics. Such an approach has potential to be used as part of a portable setup for remote and onsite analysis. After pre-processing, linear discriminant analysis indicated that this experimental setup had high discrimination efficiency (> 98% accuracies) in all studies involving the type of peppers, the origins of black peppers, and the origins of white peppers. This positive outcome suggests that this method can be applied to many more cases where whole agricultural products are directly analyzed for rapid discrimination of origins and without any sample pretreatment or workup.

## Declarations

## Acknowledgements

This work was supported by the Second Century Fund (C2F), Chulalongkorn University (for PC), and the Development and Promotion of Science and Technology Talented Project (DPST), Thailand (for MR). It was also partly supported by Thailand Research Fund (MRG 6280238), and Thai Higher Education Partnership Programme from British Council and the Ministry of Higher Education, Science, Research and Innovation, Thailand.

## Author Contributions

PC and MR conducted all experiments except chemometric analysis. NS conducted chemometric analysis. BS conducted some experiments and assisted in the operation of mass spectrometry. KW designed and supervised NS in all chemometric analysis. SM designed the data processing protocol and supervised the operation of mass spectrometry. TP designed and directed the overall flow of the study, and wrote the manuscript.

## Additional Information

A set of supplementary data is available online including an overview scheme of data preparation process, Preliminary PS-MS data, Representative MS spectra for all samples, Prediction strength values for all studies, Classification rates, and Raw numerical data in a spreadsheet.

## Competing Interest

The authors declare no competing interest.

## References

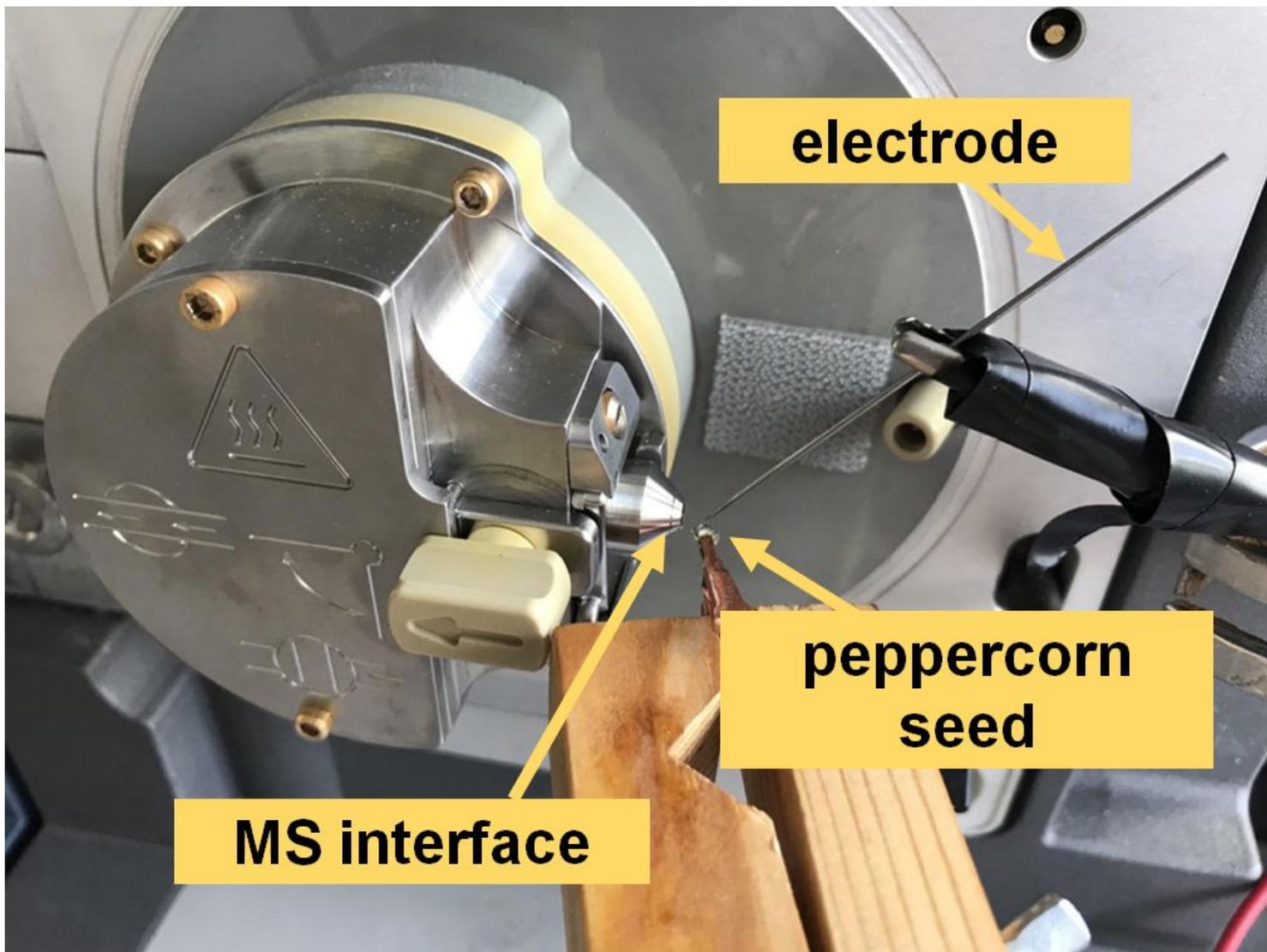
1. Maher, S., Jjunju, F. P. M. & Taylor, S. Colloquium: 100 years of mass spectrometry: Perspectives and future trends. *Rev. Mod. Phys.* **87**, 113–135 (2015).
2. Smith, B. L. *et al.* Rapid scotch whisky analysis and authentication using desorption atmospheric pressure chemical ionisation mass spectrometry. *Sci. Rep.* **9**, 7994 <https://doi.org/10.1038/s41598-019-44456-0> (2019).
3. Jeng, J. Y. *et al.* Obtaining molecular imagings of pesticide residues on strawberry surfaces with probe sampling followed by ambient ionization mass spectrometric analysis. *J. Mass Spectrom.* **56**, e4644 <https://doi.org/10.1002/jms.4644> (2020).
4. Chen, S. *et al.* Rapid analysis of bisphenol a and its analogues in food packaging products by paper spray ionization mass spectrometry. *J. Agric. Food Chem.* **65**, 4859–4865 (2017).
5. Feider, C. L., Krieger, A., DeHoog, R. J. & Eberlin, L. S. Ambient ionization mass spectrometry: Recent developments and applications. *Anal. Chem.* **91**, 4266–4290 (2019).
6. Damon, D. E. *et al.* 2d wax-printed paper substrates with extended solvent supply capabilities allow enhanced ion signal in paper spray ionization. *Analyst.* **141**, 3866–3873 (2016).

7. Suraritdechachai, S. *et al.* Rapid detection of the antibiotic sulfamethazine in pig body fluids by paper spray mass spectrometry. *J. Agric. Food Chem.* **67**, 3055–3061 (2019).
8. Damon, D. E. *et al.* Determining surface energy of porous substrates by spray ionization. *Langmuir.* **35**, 13853–13859 (2019).
9. Sarih, N. M. *et al.* Accelerated nucleophilic substitution reactions of dansyl chloride with aniline under ambient conditions via dual-tip reactive paper spray. *Sci. Rep.* **10**, 21504 <https://doi.org/10.1038/s41598-020-78133-4> (2020).
10. Liu, J. *et al.* Development, characterization, and application of paper spray ionization. *Anal. Chem.* **82**, 2463–2471 (2010).
11. Gross, J. H. Direct analysis in real time—a critical review on dart-ms. *Anal. Bioanal. Chem.* **406**, 63–80 (2014).
12. Marić, M., Marano, J., Cody, R. B. & Bridge, C. Dart-ms: A new analytical technique for forensic paint analysis. *Anal. Chem.* **90**, 6877–6884 (2018).
13. Forbes, T. P., Sisco, E., Staymates, M. & Gillen, G. Dart-ms analysis of inorganic explosives using high temperature thermal desorption. *Anal. Methods.* **9**, 4988–4996 (2017).
14. Ifa, D. R., Wu, C., Ouyang, Z. & Cooks, R. G. Desorption electrospray ionization and other ambient ionization methods: Current progress and preview. *Analyst.* **135**, 669–681 (2010).
15. Garza, K. Y. *et al.* Desorption electrospray ionization mass spectrometry imaging of proteins directly from biological tissue sections. *Anal. Chem.* **90**, 7785–7789 (2018).
16. Smith, M. J. P., Cameron, N. R. & Mosely, J. A. Evaluating atmospheric pressure solids analysis probe (asap) mass spectrometry for the analysis of low molecular weight synthetic polymers. *Analyst.* **137**, 4524–4530 (2012).
17. Tose, L. V., Murgu, M., Vaz, B. G. & Romão, W. Application of atmospheric solids analysis probe mass spectrometry (asap-ms) in petroleomics: Analysis of condensed aromatics standards, crude oil, and paraffinic fraction. *J. Am. Soc. Mass Spectrom.* **28**, 2401–2407 (2017).
18. Jjunju, F. P. M. *et al.* Analysis of polycyclic aromatic hydrocarbons using desorption atmospheric pressure chemical ionization coupled to a portable mass spectrometer. *J. Am. Soc. Mass Spectrom.* **26**, 271–280 (2015).
19. Jjunju, F. P. M. *et al.* Hand-held portable desorption atmospheric pressure chemical ionization ion source for in situ analysis of nitroaromatic explosives. *Anal. Chem.* **87**, 10047–10055 (2015).
20. Jafari, M. T., Khayamian, T., Shaer, V. & Zarei, N. Determination of veterinary drug residues in chicken meat using corona discharge ion mobility spectrometry. *Anal. Chim. Acta.* **581**, 147–153 (2007).
21. Smith, B. L. *et al.* Flexible drift tube for high resolution ion mobility spectrometry (flex-dt-ims). *Anal. Chem.* **92**, 9104–9112 (2020).
22. Song, L., You, Y., Perdomo, N. R. & Evans-Nguyen, T. Inexpensive ultrasonic nebulization coupled with direct current corona discharge ionization mass spectrometry for liquid samples and its fundamental investigations. *Anal. Chem.* **92**, 11072–11079 (2020).

23. Mullen, M. & Giordano, B. C. Combined secondary electrospray and corona discharge ionization (secdi) for improved detection of explosive vapors using drift tube ion mobility spectrometry. *Talanta*. **209**, 120544 <https://doi.org/10.1016/j.talanta.2019.120544> (2020).
24. Sekimoto, K. *et al.* Improvement in ionization efficiency of direct analysis in real time-mass spectrometry (dart-ms) by corona discharge. *Analyst*. **141**, 4879–4892 (2016).
25. Chingin, K. *et al.* Rapid detection of mycobacterium tuberculosis cultures by direct ambient corona discharge ionization mass spectrometry of volatile metabolites. *RSC Adv.* **6**, 59749–59752 (2016).
26. Hu, L. *et al.* Early release of 1-pyrroline by pseudomonas aeruginosa cultures discovered using ambient corona discharge ionization mass spectrometry. *RSC Adv.* **6**, 8449–8455 (2016).
27. FAO. *Faostat*, <http://www.fao.org/faostat/en/#data/TP> (2021).
28. Wang, P. & Yu, Z. Species authentication and geographical origin discrimination of herbal medicines by near infrared spectroscopy: A review. *J. Pharm. Anal.* **5**, 277–284 (2015).
29. Lawless, L. J. R., Hottenstein, A. & Ellingsworth, J. The mccormick spice wheel: A systematic and visual approach to sensory lexicon development. *J. Sens. Stud.* **27**, 37–47 (2012).
30. Witkowska, A. M., Hickey, D. K., Alonso-Gomez, M. & Wilkinson, M. G. The microbiological quality of commercial herb and spice preparations used in the formulation of a chicken supreme ready meal and microbial survival following a simulated industrial heating process. *Food Control.* **22**, 616–625 (2011).
31. Tripathy, V., Basak, B. B., Varghese, T. S. & Saha, A. Residues and contaminants in medicinal herbs—a review. *Phytochem. Lett.* **14**, 67–78 (2015).
32. Wilde, A. S., Haughey, S. A., Galvin-King, P. & Elliott, C. T. The feasibility of applying nir and ft-ir fingerprinting to detect adulteration in black pepper. *Food Control.* **100**, 1–7 (2019).
33. Lafeuille, J. L., Frégière-Salomon, A., Michelet, A. & Henry, K. L. A rapid non-targeted method for detecting the adulteration of black pepper with a broad range of endogenous and exogenous material at economically motivating levels using micro-atr-ft-mir imaging. *J. Agric. Food Chem.* **68**, 390–401 (2020).
34. Josling, T. The war on terroir: Geographical indications as a transatlantic trade conflict. *J. Agric. Econ.* **57**, 337–363 (2006).
35. WTO. *Intellectual property (trips) - geographical indications - background and the current situation*, [https://www.wto.org/english/tratop\\_e/trips\\_e/gi\\_background\\_e.htm](https://www.wto.org/english/tratop_e/trips_e/gi_background_e.htm) (2021).
36. EU. *Geographical indications—a european treasure worth €75 bn*, [https://ec.europa.eu/commission/presscorner/detail/en/IP\\_20\\_683](https://ec.europa.eu/commission/presscorner/detail/en/IP_20_683) (2021).
37. Luykx, D. M. A. M. & van Ruth, S. M. An overview of analytical methods for determining the geographical origin of food products. *Food Chem.* **107**, 897–911 (2008).
38. Dias, C. & Mendes, L. Protected designation of origin (pdo), protected geographical indication (pgi) and traditional speciality guaranteed (tsg): A bibliometric analysis. *Food Res. Int.* **103**, 492–508 (2018).

39. Consonni, R. & Cagliani, L. R. in *Advances in food and nutrition research* Vol. 59 (ed Steve L. Taylor)87–165(Academic Press, 2010).
40. Brereton, R. G. *Chemometrics: Data analysis for the laboratory and chemical plant* 1st edn (Wiley, 2003).
41. Brereton, R. G. *Applied chemometrics for scientists* 1st edn (Wiley, 2007).
42. Brereton, R. G. *Chemometrics for pattern recognition* 1st edn (Wiley, 2009).
43. Wongravee, K. *et al.* Variable selection using iterative reformulation of training set models for discrimination of samples: Application to gas chromatography/mass spectrometry of mouse urinary metabolites. *Anal. Chem.* **81**, 5204–5217 (2009).
44. Ravindran, P. N. & Kallapurackal, J. A. in *Handbook of herbs and spices* (ed K. V. Peter)86–115(Woodhead Publishing, 2012).
45. Singh, G., Marimuthu, P., Catalan, C. & deLampasona, M. Chemical, antioxidant and antifungal activities of volatile oil of black pepper and its acetone extract. *J. Sci. Food Agric.* **84**, 1878–1884 (2004).
46. Jirovetz, L., Buchbauer, G., Ngassoum, M. B. & Geissler, M. Aroma compound analysis of piper nigrum and piper guineense essential oils from cameroon using solid-phase microextraction–gas chromatography, solid-phase microextraction–gas chromatography–mass spectrometry and olfactometry. *J. Chromatogr. A.* **976**, 265–275 (2002).
47. Kapoor, I. P. S. *et al.* Chemistry and in vitro antioxidant activity of volatile oil and oleoresins of black pepper (piper nigrum). *J. Agric. Food Chem.* **57**, 5358–5364 (2009).
48. Li, Y. *et al.* Analysis of chemical components and biological activities of essential oils from black and white pepper (piper nigrum L.) in five provinces of southern china. *LWT.* **117**, 108644 <https://doi.org/10.1016/j.lwt.2019.108644> (2020).
49. Wongravee, K. *et al.* Monte-carlo methods for determining optimal number of significant variables. Application to mouse urinary profiles. *Metabolomics.* **5**, 387 <https://doi.org/10.1007/s11306-009-0164-4> (2009).
50. Esslinger, S., Riedl, J. & Fauhl-Hassek, C. Potential and limitations of non-targeted fingerprinting for authentication of food in official control. *Food Res. Int.* **60**, 189–204 (2014).

## Figures



**Figure 1**

Experimental setup for direct ionization MS on peppercorn seeds.

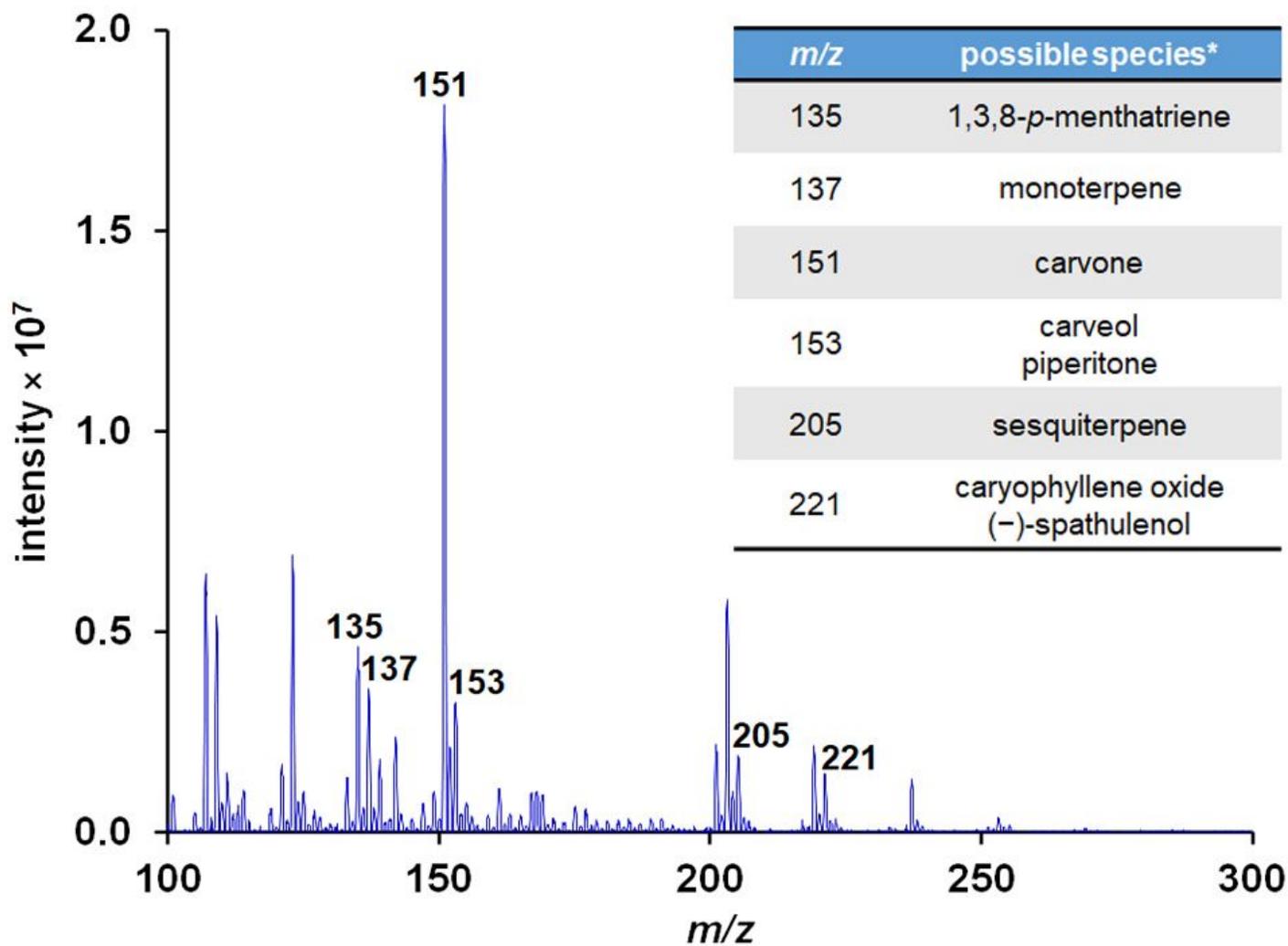
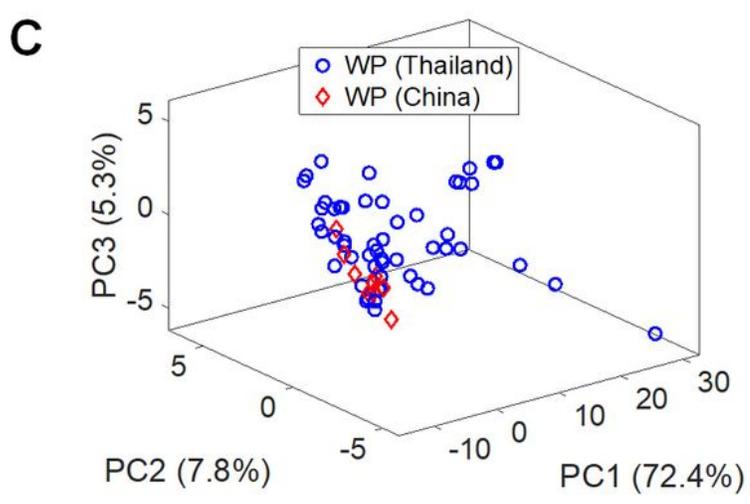
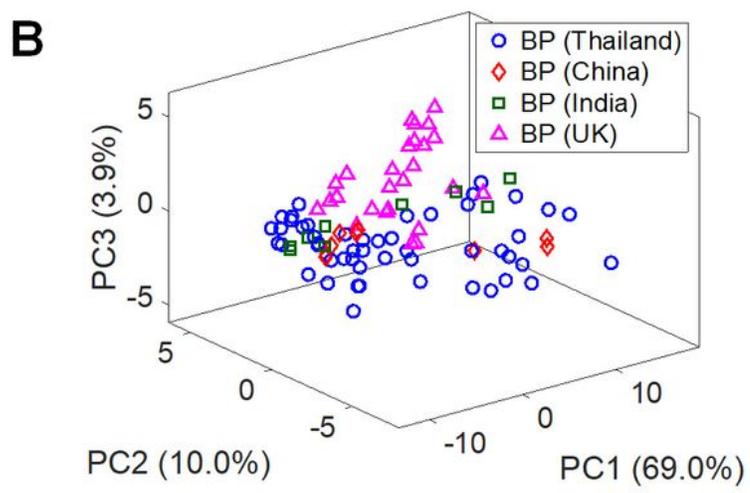
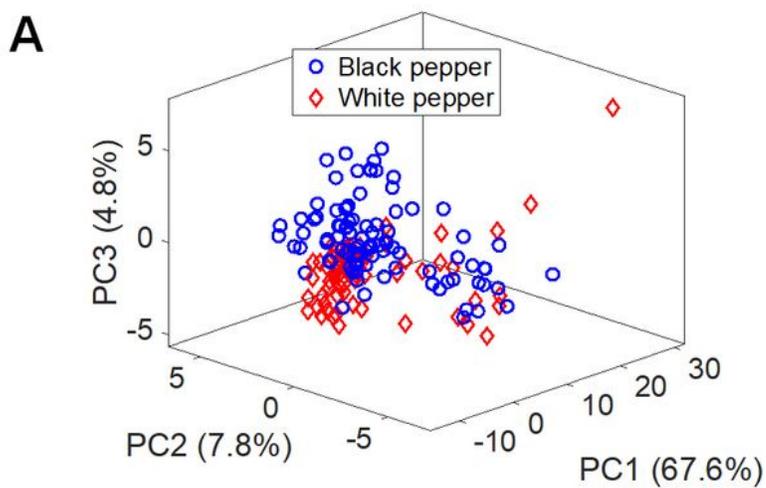


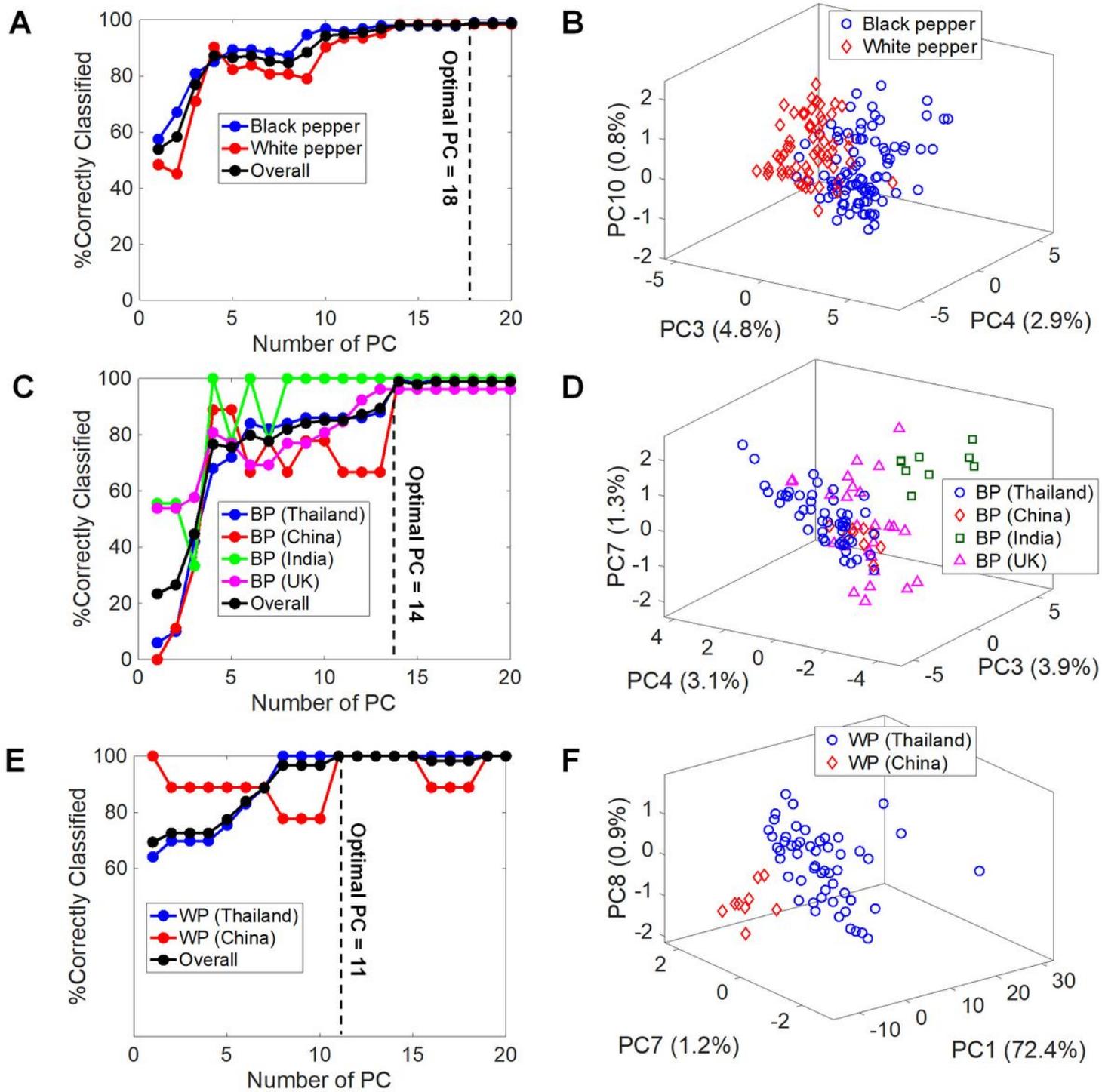
Figure 2

A representative MS spectrum from the sample BP(Thailand) – 06, along with some examples of putative chemical species that were deduced from the literature<sup>45-48</sup>.



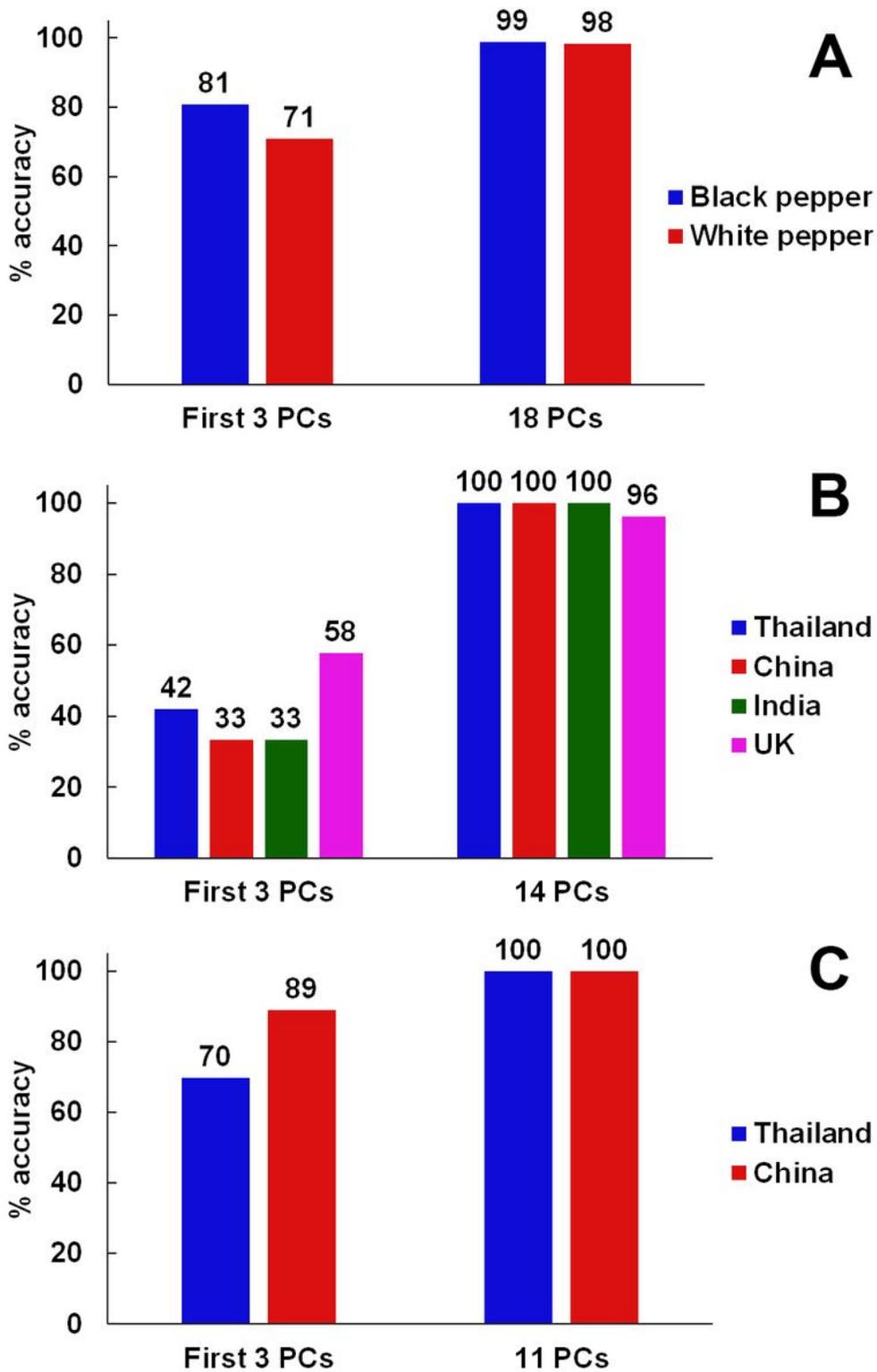
**Figure 3**

PC score plots of the first 3 principal components of the pre-processed data using all m/z data points to visualize the cluster relationship of (A) types of peppers, (B) origins of black peppers and (C) origins of white peppers.



**Figure 4**

The percentage of correctly classified samples with the number of PCs used to build a model (PC1-20), along with their respective PC score plots using the best discriminate PCs for the study of (A) the types of peppers, (B) is the corresponding PC score plot, (C) the origins of black peppers, (D) is the corresponding PC score plot, and (E) the origins of white peppers with (F) showing the corresponding PC score plot.



**Figure 5**

Classification accuracies of A) the discrimination between black and white peppers, B) the geographical discrimination of black peppers, and C) the geographical discrimination of white peppers.

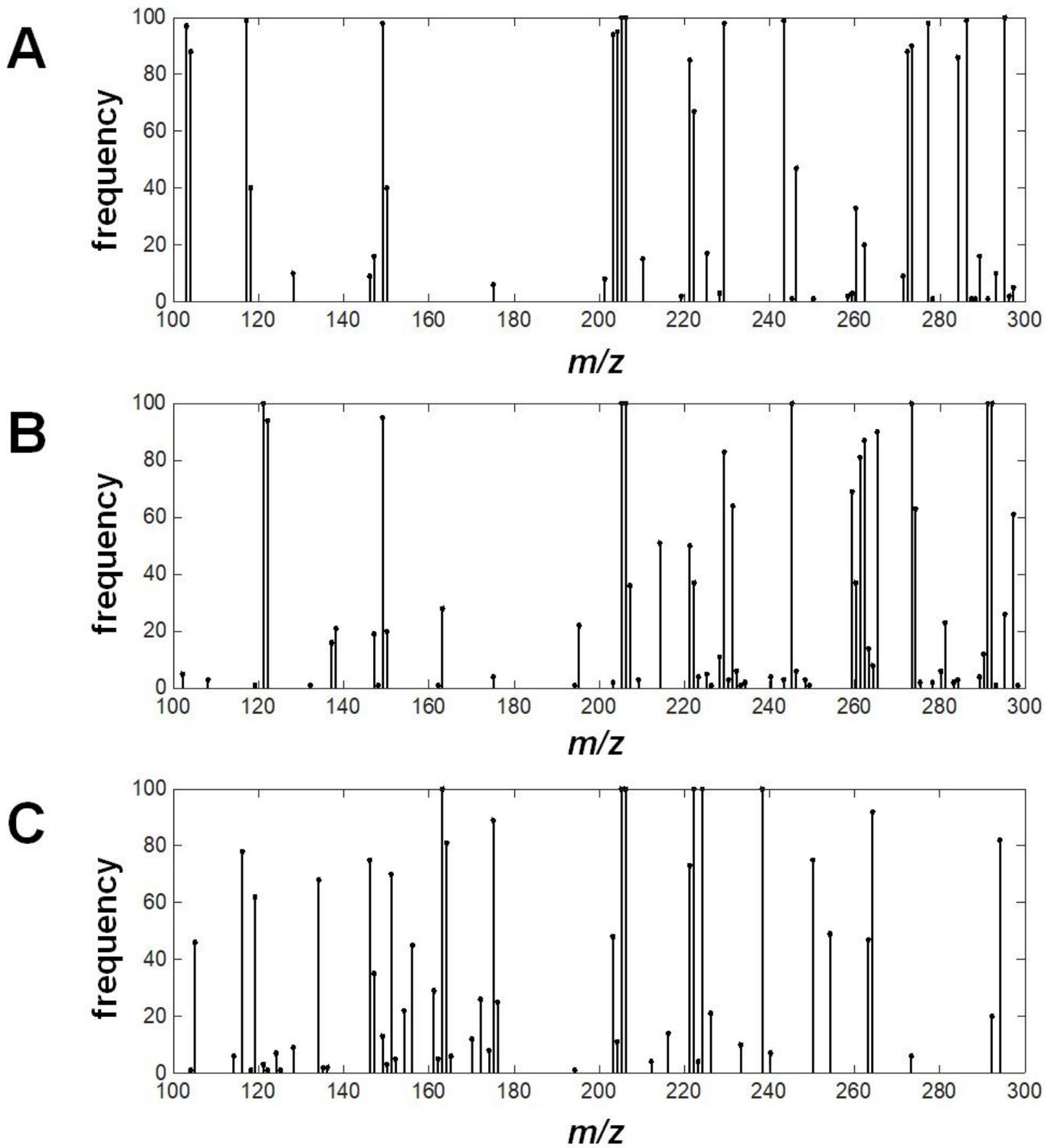


Figure 6

Bar charts showing the number of times each variable ( $m/z$  data points) was found in 100 iterations.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [RawDataforPublication.xlsx](#)
- [SupplInfo.02.pdf](#)