Measuring Plant Metabolite Abundance in Spearmint (*Mentha spicata* L.) with Raman Spectra to Determine Optimal Harvest Time

Jingzhe Li[ab], Chamari S. Wijesooriya[a], Sadie J. Burkhow[ab], Linda K.B. Brown[c], Beatrice Y. Collet[c], John A. Greaves[c], Emily A. Smith[ab*]

[a]Department of Chemistry

Iowa State University

Ames IA 50011, United States

[b]Chemical and Biological Science Division

The Ames Laboratory, United States, Department of Energy

311 Iowa State University, Ames, IA, 50011, United States

[c]Kemin Industries, Inc

1900 Scott Avenue

Des Moines IA 50317, United States

*Telephone: 515-294-1424, Email: esmith1@iastate.edu

Abstract:

A fast field-deployable method utilizing Raman spectroscopy to determine the optimal harvest time of plants to extract the highest abundance of target metabolites is presented. Rosmarinic acid is a metabolite extracted from spearmint (*Mentha spicata* L.). Leaves from commercial 'Native' and proprietary clonal line 'KI110' spearmint were measured as a function of cell type and age to determine rosmarinic acid abundance. A linear regression model with leave-one-out cross-validation (R^2_{CV} = 0.61, RMSECV = 11.1 mg/g) was developed between selected Raman peak areas and rosmarinic acid concentrations determined by high-performance liquid chromatography (HPLC). A principal component analysis (PCA) model was also developed to determine rosmarinic acid abundance. The method may be suited to the analysis of many agriculturally-relevant plant species and metabolites with distinct Raman peaks.

Keywords:

Raman Spectroscopy, *in situ* Plant Metabolite Quantification, Vibrational Spectroscopy, Rosmarinic Acid, Plant Harvest, Plant Leaves, Linear Regression, Principal Component Analysis

1. Introduction:

There is a need for rapid analysis methods to confirm the optimal plant harvest time in order to extract metabolites found within the tissues. Primary and secondary plant metabolites include amino acids, lipids, flavonoids, polyphenols, alkaloids, nitriles, and chlorophylls, to name a few. Unlike harvesting fruits or a flowering plant, where a visual cue is easily observed and may be suitable for determining the optimal harvest time, there is often no visual cue when harvesting a vegetative plant, such as spearmint, for a target metabolic component, such as rosmarinic acid. Typical methods to investigate the plant metabolite concentration include extraction from the tissue followed by high-performance liquid chromatography (HPLC) quantification. Such an analysis cannot be performed in the field, requires lengthy and labor-intensive liquid sample preparation, and often requires abundant tissue (on the order of grams) for the analysis. Metabolite expression and accumulation is dependent on many biotic and abiotic factors, including temperature, humidity, soil type, and amount of sunlight.[1-2] As a result, the best harvest time also depends on many factors and is not necessarily predictable. Thus, an alternative rapid analysis method to determine the plant metabolite concentration for optimal harvest time is needed.

Raman spectroscopy could be an alternative technique to determine promptly when a plant is ready to harvest in order to obtain the most extractable metabolite(s). As a non-destructive analytical technique that provides information on molecular functional groups, Raman spectroscopy can be used to differentiate many plant metabolites by characteristic vibrational peaks.[3-4] The sample preparation for Raman spectroscopy is usually minimal; analyses can often be performed on whole leaves or other tissues. A spectrum can be acquired within seconds in many cases and directly interpreted or screened with spectral libraries. Also, handheld commercial Raman spectrometers can be carried directly to a greenhouse or field and used with minimal training.[5]

Quantitation with Raman spectroscopy is possible because the intensity of the Raman scattered light from the sample, $I(v)_R$ is proportional to the number of molecules producing the scattered light, N.[6] In other words, $I(v)_R \propto \mathrm{N}$. Several research groups have demonstrated the potential of using Raman spectroscopy and univariate calibration with a calibration curve for real-time monitoring of lignocellulosic bioethanol production, including pretreatment, hydrolysis, and fermentation steps.[7-12] However, the complexity of biomass and biomass-derived samples often limits the utility of univariate calibration with these types of samples. Another direction for quantitative analysis is to use Raman spectra along with data from another analysis method (such as HPLC) to build regression models. In this way, each method's instrument response can be directly correlated to the metabolite concentration using regression-based analysis. Combined with regression analysis, Raman spectroscopy has been used to determine the concentration of plant metabolites.[13-15] Anastasaki and coworkers determined that Raman spectroscopy could be a rapid screening tool for saffron (crocetin esters) quality.[14] They used HPLC and UV/visible spectrophotometry to measure concentrations of the crocetin esters and correlated those data with Raman spectra of the same materials using partial least squares (PLS) regression. Raman spectroscopy has been combined with chemometric techniques, such as principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA), to establish classification models in the field of agriculture and food science. These classifications include rapid discrimination of plant variety, species and maturity, as well as food quality control.[16-21]

Rosmarinic acid, a polyphenol plant metabolite (Supplemental Figure 1), has many uses in support of human health when taken as a nutritional supplement.[22] It is most notably used as an antioxidant,[23] anti-inflammatory,[24] anti-mutagen,[25] and an astringent.[26] Rosmarinic acid is commonly extracted from the *Lamiaceae* (i.e., mint or sage) and *Boraginaceae* plant families.[27] Sik *et al.* discussed various extraction methods for rosmarinic acid from multiple plants.[27] There is considerable variation in the amount of extractable rosmarinic acid in spearmint plants.[28] For

example, plant maturity leads to an accumulation of rosmarinic acid. Rosmarinic acid is found within the roots, stems, and leaves of plants that synthesize this metabolite.[28] Leaves have the highest concentration of rosmarinic acid, followed by the stems and then the roots.[29] Leaves are generally the most relevant source of rosmarinic acid in the commercial harvest of this metabolite.

In this work, data are presented showing that Raman spectroscopy is a rapid, non-invasive technique to monitor rosmarinic acid abundance for the determination of optimal spearmint harvest time. We focused on rosmarinic acid in 'KI110'[30-31] and commercial 'Native' spearmint plants. Raman data were analyzed with regression analysis, statistical analysis as well as principal component analysis for the rapid and non-destructive determination of the rosmarinic acid abundance in spearmint plants.

2. Materials and Methods:

2.1 Materials

Native and KI110 spearmint (*Mentha spicata* L.) plants were used in this study. The KI110 clonal line was developed to accumulate rosmarinic acid up to 100 mg/g on a dry weight basis, whereas the commercial Native spearmint plant accumulates <20 mg/g in leaf tissue under field conditions in early June.[31] The plants were vegetatively propagated to produce multiple genetically identical clones in 4-inch pots. These pots were grown at the Kemin Specialty Crop Improvement Research Farm near Kelley, Iowa, as well as inside Kemin's greenhouses in Des Moines, Iowa. The plants were harvested at approximately three months of age from August to October 2019 and were also provided in 4-inch pots for measurements.

Nine sets of pots were measured (labeled I to IX). Within each set, there were varying numbers of pots that were measured (Supplemental Table 1). Sets I to IV consisted of the following numbers of KI110 pots: 5, 7, 4, and 5, respectively. The remaining sets V to IX consisted

of the following numbers of Native spearmint pots 2, 3, 2, 7, and 7, respectively. Sets IV, VII, VIII, and IX were grown in the greenhouse, and all other sets were grown in field plots.

2.2 High-Performance Liquid Chromatography (HPLC)

For HPLC analysis of each set of spearmint plants, the dried leaf mass of all pots from a set was combined to extract the rosmarinic acid. The rosmarinic acid extraction process used a KH_2PO_4 buffer:reagent alcohol (1:1 v/v) extraction solution. The reagent alcohol is 90% ethyl alcohol denatured with 5% isopropanol and 5% methanol. An HPLC instrument (Agilent 1100) with a Quad pump, an autosampler, a column thermostat maintained at 35 °C and a diode array detector were used. A C18 column was used and the chromatographic mobile phase ramping scheme is shown in Supplemental Table 2. All measurements were collected using a 5 µL injection volume and a flow rate of 1 mL/min.

A stock solution was made of 0.8652 mg/mL rosmarinic acid in acetonitrile. The solutions with the following rosmarinic acid concentrations were then made by dilution with acetonitrile: 0.2163, 0.4326, 0.6589, 0.7354, and 0.8651 mg/mL. The calibration curve that was used to determine the rosmarinic acid concentration was generated with the peak eluting at ~5 minutes. Rosmarinic acid concentrations measured by HPLC on a dry weight basis (mg/g) are shown in Table 1.

2.3 Raman Measurements

The Raman spectra were measured with an XploRA Plus Raman confocal upright microscope (HORIBA Scientific, Edison, New Jersey) with a Synapse EMCCD camera. An irradiance of 17 kW/cm^2 with a 785-nm solid-state laser was used. An Olympus objective (20× magnification, 0.4 numerical aperture) was used to collect optical images and Raman data in the epi-direction. Raman measurements were collected with a 600 grooves/millimeter (resolution ± 10 cm^{-1}) grating with a 15s acquisition and 4 accumulations. The analyzed spectral range was

from 300 cm^{-1} to 1800 cm^{-1}. A 300 μm monochromator slit and 100 μm confocal pinhole were used for data collection. White-light optical images for representative cell areas were collected and analyzed with the imaging analysis software ImageJ. The Raman instrument used in this study required the leaves to be positioned flat. Leaves were cut from the plant and placed with the adaxial leaf surface facing the air on a coverslip coated with double-sided tape to ensure it laid flat, and were measured immediately.

Leaves were the sole focus of this study to mimic what is most important to the commercial harvest of rosmarinic acid. In total, 633 leaves were measured and 3,756 Raman spectra were collected for this study to capture a large amount of variance. For each pot, leaves of different ages were measured: leaf 3 refers to the third leaf from the top of the stalk, leaf 4 is the fourth from the top, and leaf 5 is fifth from the top. Leaves emerge from the top of the plant, so the age is leaf 3 < leaf 4 < leaf 5. A total of 15 to 18 leaves were measured per pot. For each leaf, 2 to 3 spectra were collected from different locations on the leaf where epidermal cells were present to represent a larger surface area of the leaf. Each spectrum was background subtracted[32] and then the 2 to 3 background-subtracted spectra were averaged for each leaf (i.e., the idea is that the average of many locations represents a larger area). This procedure was repeated for areas of the leaf where trichomes were present. For each leaf, this resulted in 1 Raman spectrum for the epidermal cells and 1 Raman spectrum for the trichomes. Since the entire plant mass is harvested at the same time, a final averaging step was then performed. The Raman spectra from 3 leaves (of the same leaf age and cell type from the same spearmint plant) were averaged. Thus, the 3,756 Raman spectra were reduced to 422 representative spectra. Throughout the text, these are referred to as representative Raman spectra; whereas the average Raman spectra refer to the 9 spectra for sets I-IX that are shown in Figure 1 and are averaged from all the spectra collected for a given set).

The average Raman spectrum of a rosmarinic acid solution with 3 replicate measurements was collected as a control (Figure 1A). The aqueous solution (10 mg/ml) was prepared from rosmarinic acid powder (Sigma-Aldrich) and DI water.

2.4 Data and Statistical Analysis

2.4.1 Raman Peak Amplitude and Area

MATLAB 2019b (MathWorks, Natick, MA, USA) was used for all data analysis unless otherwise noted. For all 422 representative Raman spectra, the 1600 cm^{-1} peak, which is partially assigned to rosmarinic acid, was fit with a Lorentzian function[33] with a linear baseline to extract the peak amplitude and peak area. Other Raman peaks were similarly fit to obtain the values needed to calculate Raman peak area ratios. The peak amplitude of all fitted peaks was larger than the limit of detection, which was determined by three times of the standard deviation of the spectral region from 2000 to 2025 cm^{-1}, where no Raman peaks were present (i.e., where there is spectral noise).

2.4.2 Linear Regression Analysis

For the linear regression analysis, the average 1600 cm^{-1} Raman peak area from all representative Raman spectra in a set was calculated. The linear regression model was built with leave-one-out cross-validation with the nine Raman peak areas and the nine rosmarinic acid concentrations determined by HPLC, which were averaged for all nine sets of plants. The model was evaluated by R^2_{CV} and RMSECV calculated with the following equations:

$$R^2_{CV} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}$$

where n represents the number of samples, y_i represent the *i*th experimental rosmarinic acid concentration determined by HPLC, $\hat{y}_i$ represents the *i*th rosmarinic acid concentration predicted by using the model obtained without using the *i*th sample during cross-validation, $\bar{y}$ represent the average of the experimental rosmarinic acid concentrations.

2.4.3 Statistical Analysis

A statistically significant difference (i.e., *p*-value < 0.0001) *among two data sets* was determined using the unpaired two-tailed t-test with equal or unequal variance (as determined by the two tailed F-test). A statistically significant difference (i.e., *p*-value < 0.05) *among more than two data sets* was determined with a one-way analysis of variance (ANOVA) using Microsoft Excel. The confidence interval was 95% and the alpha (α) value was 0.05. Supplemental Table 3 lists *p*-values for all the statistical analyses.

2.4.4 Chemometric Analysis

All 422 representative Raman spectra were preprocessed by standard normal variate transformation[34-35] prior to the chemometric analysis. Principle component analysis (PCA) was conducted for all preprocessed Raman spectra by the 'pca()' function in MATLAB with a singular value decomposition algorithm as the default. The Raman data were decomposed into two orthogonal matrices, the score matrix and the loading matrix; these explain the variance in the Raman data and were used to generate the score plot and PCA loading plots.

3. Results and Discussion:

3.1 Raman Spectra and Peak Assignments of Rosmarinic Acid in Spearmint Plants

The KI110 plants produce a higher concentration of rosmarinic acid and using both Native and KI110 plants enabled a larger range of rosmarinic acid concentrations to be included in this study. Average Raman spectra for each of the 9 sets of spearmint plants are displayed in Figure 1A. The peak assignments are shown in Supplemental Table 4. The primary difference between the KI110 and Native plant spectra is the 1600 ± 2 cm^{-1} (n = 422) peak area. In the Raman spectrum of a rosmarinic acid solution (Figure 1A), the most intense peak is at ~1600 cm^{-1}, which is consistent with the value of 1597 cm^{-1} reported in the literature.[33] For rosmarinic acid solutions, there is a linear increase in the 1600 cm^{-1} peak area with increasing rosmarinic acid concentrations (data not shown). It is assumed that the peak area difference measured for the KI110 and Native plants reflects, in part, differences in rosmarinic acid concentration. In plant tissues, the 1600 cm^{-1} peak is often assigned to aromatic species, including the phenylpropanoid-based lignin polymers.[36] Lignin is not typically prevalent in the cell wall of plant leaves (although it is found in the stem), however, other aromatic species may be present. A variety of essential oils are commonly found within spearmint. The most abundant essential oil in Native spearmint is carvone,[37] which has a Raman peak at 1623 cm^{-1} that is assigned to the ring C=C stretching vibrations.[38] The percentage of carvone relative to all extracted essential oils in Native spearmint grown in India and Illinois were 46.9-76.6%[39] and 70.9%,[40] respectively. The cultivars used within the current study were bred to have minimal carvone (data not shown).

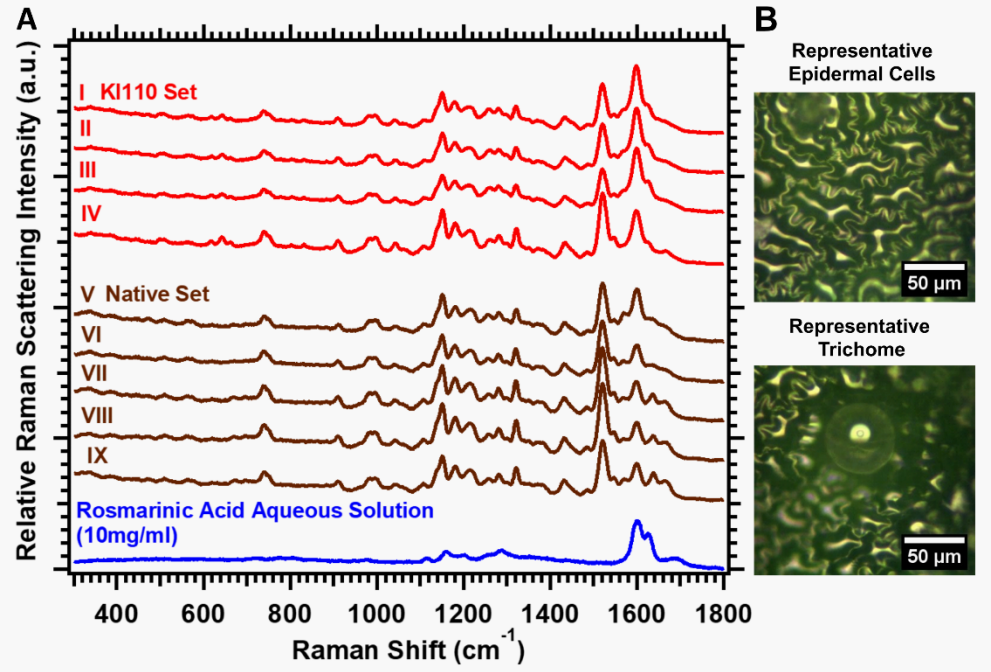

Figure 1. (A) Average Raman spectra of KI110 sets (red), Native sets (brown), and an aqueous solution of rosmarinic acid (blue). The number of spectra that were averaged for each set is listed in Supplemental Table 1. The Raman peak at 1600 cm^{-1} is assigned, in part, to rosmarinic acid. The spectra have been offset and background subtracted for clarity. (B) Representative optical images of epidermal cells and a trichome (also showing some epidermal cells) on an adaxial leaf surface where Raman measurements were collected. The spectra shown in panel A were averages of both epidermal cells and trichomes.

3.2 Correlation of the 1600 cm^{-1} Peak Area to Rosmarinic Acid Abundance

It is common to measure ratios of Raman peak areas when differences in the signal intensity may be caused by differences in sample thickness or surface roughness; this may be the case with leaves. The peak area representing a component with varying abundance and a peak area representing a non-varying component are ratioed. Several Raman peak area ratios were analyzed to check their correlation with a simple linear regression to the rosmarinic acid

concentration measured with HPLC; the 1600 cm^{-1} peak areas were also used in a regression analysis. Table 1 shows the ratios for the 1600 cm^{-1} peak area relative to the 740, 1000, and 1521 cm^{-1} peak area. The 740 and 1521 cm^{-1} peaks were primarily assigned to chlorophyll[41] and carotenoids,[3] respectively. The 1000 cm^{-1} peak was assigned to both cellulose and carotenoids.[3, 42] The 1600 cm^{-1} Raman peak areas showed the best correlation to rosmarinic acid concentration with an R^2 value of 0.71; thus the 1600 cm^{-1} Raman peak areas were used for all further analyses. A linear regression model between the 1600 cm^{-1} Raman peak areas and HPLC concentrations for all nine sets of data was established. The R^2_{CV} value was 0.61 and RMSECV was 11.1 mg/g for leave-one-out cross-validation.

While the regression coefficient may seem low, it is important to consider the many differences between the Raman measurement and the HPLC measurement. Raman spectroscopy reflects the abundance of rosmarinic acid within the leaf, which is not necessarily the amount of rosmarinic acid available for extraction. Extraction is rarely 100% efficient. In addition, the Raman analysis likely measures only the rosmarinic acid near the adaxial surface since absorption and scattering limit the light penetration through the leaf. While the HPLC analysis directly measures the extracted rosmarinic acid, it is time and cost consuming to perform. Compared with hours to days to obtain HPLC data, it only takes minutes to complete the Raman analysis and this may facilitate large-scale daily or weekly measurements of many plants.

Table 1. Raman peak area ratios, 1600 cm^{-1} peak areas, and HPLC-measured rosmarinic acid concentrations.

	(1600 cm^{-1} /740 cm^{-1}) peak area ratios[a]	(1600 cm^{-1} /1000 cm^{-1}) peak area ratios[a]	(1600 cm^{-1} /1521 cm^{-1}) peak area ratios[a]	1600 cm^{-1} Raman peak areas[a]	HPLC Rosmarinic acid concentration (mg/g)[b]
	Avg. ± Std.	Avg. ± Std.	Avg. ± Std.	Avg. ± Std. × 10^4	Avg. ± Std.
KI110 Set					
I	20 ± 20	30 ± 20	3 ± 2	22 ± 9	45.6 ± 0.6
II	20 ± 20	40 ± 20	3 ± 2	21 ± 9	50 ± 1
III	25 ± 3	47 ± 8	3.7 ± 0.8	21 ± 8	41.0 ± 0.2
IV	10 ± 3	28 ± 9	1.6 ± 0.6	18 ± 7	47 ± 1
Native Set					
V	20 ± 10	30 ± 20	2 ± 2	17 ± 9	6.8 ± 0.8
VI	10 ± 6	20 ± 20	1.4 ± 0.8	12 ± 5	6.1 ± 0.6
VII	5 ± 3	17 ± 8	1.1 ± 0.8	12 ± 8	13.1 ± 0.2
VIII	6 ± 3	18 ± 9	1.0 ± 0.6	12 ± 7	12.5 ± 0.6
IX	6 ± 4	20 ± 10	1.3 ± 0.9	12 ± 8	15.8 ± 0.6
R^2 [c]	0.42	0.57	0.47	0.71	--

[a] Averages from all representative spectra for all plants within sets I, II, III, *etc*., including epidermal cells and trichomes.
[b] On a dry weight basis. The sample mass from all pots within each set was collected and divided into three equal parts for HPLC measurements. Average value from three replicate measurements.
[c] R^2 is from the linear correlation between the peak area ratios or peak areas and the mean rosmarinic acid concentration measured by HPLC. (The correlation analysis was performed on the values prior to rounding to the correct number of significant figures).

3.3 Rosmarinic Acid Abundance: Cellular Structure and Leaf Age

The statistical analyses of the 1600 cm^{-1} Raman peak areas for all 422 spectra were further evaluated to determine if differences in the plant type, cellular structure or leaf age affect rosmarinic acid abundance. Figure 2A shows the distribution of 1600 cm^{-1} Raman peak areas for leaves from KI110 and Native plants. The *p*-value using the t-test is smaller than 0.0001, which indicates a significant difference of the peak areas between the two types of plants.

There were two relevant structures within the spearmint adaxial leaf surface, epidermal cells and trichomes. Optical images of representative epidermal cells and a trichome are displayed in Figure 1B. Figure 2B shows the plot of the 1600 cm^{-1} peak area for KI110 and Native trichomes and epidermal cells. As expected, the peak areas of either the epidermal cells or

trichomes of the KI110 plants were statistically larger than the Native plants, with the *p*-value less than 0.0001. The peak areas for the trichomes were larger than for the epidermal cells for both types of plants, with the *p*-values less than 0.0001. These data indicate that rosmarinic acid abundance varies within the cellular structures of the leaves. In order of rosmarinic acid abundance: KI110 trichomes > Native trichomes > KI110 epidermal cells > Native epidermal cells. This conclusion is very useful: spatially-dependent physiological information about rosmarinic acid in spearmint leaves has been gained using Raman spectroscopy. In order to limit the influence of different cellular structures on the analysis, a large sample area should be measured when comparing different plant populations. The measured area is controlled by the size of the laser spot, or as used in this study, by averaging the spectra from many locations across the leaf. Finally, the distribution of metabolites among different cellular structures must always be considered when applying a spectroscopy technique to plant samples since the observation of different cellular structures may lead to different conclusions. Thus we considered the influence of cellular structures on all additional analyses.

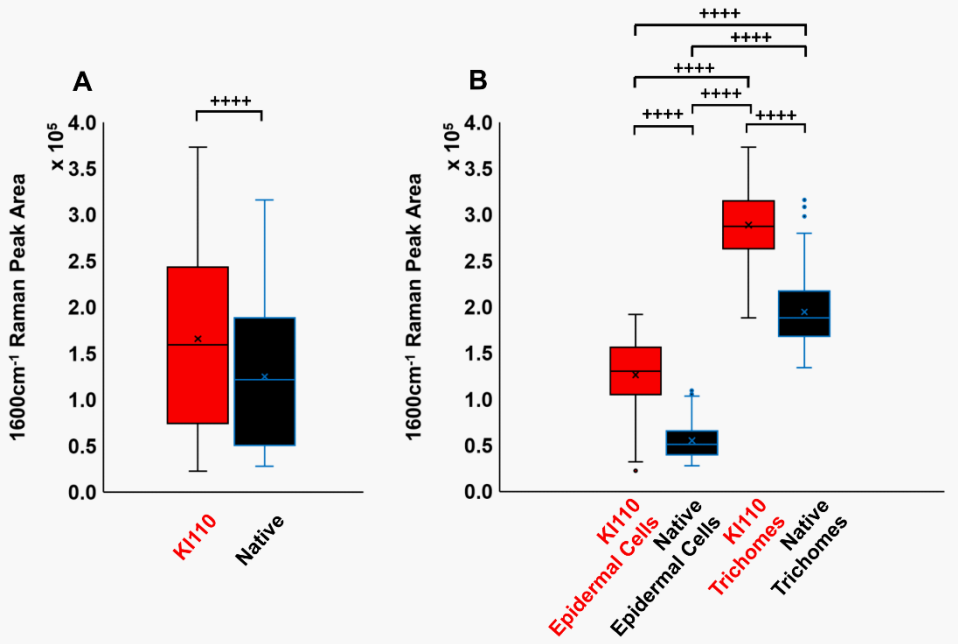

Figure 2. (A) Box and whisker plot of the rosmarinic acid abundance determined from the 1600 cm^{-1} peak area within KI110 and Native plants. (B) Box and whisker plot of the rosmarinic acid abundance determined from the 1600 cm^{-1} peak area for the epidermal cells and trichomes of KI110 and Native plants. A *p*-value < 0.0001 was represented with ++++ in both plots.

The statistical analysis of rosmarinic acid abundance within leaves of different ages was explored to see if there was variation between younger and older leaves. There were hundreds of leaves on each spearmint plant. New leaves emerge from the top of the seedling, and the leaves farther down the stem are older. In this study, the leaves were counted from the top of the plant, as the older leaves can be the first to die and may not be consistent from plant to plant. Leaf 3 was the third leaf from the top of the plant stalk, and so on for leaf 4 and 5. Figure 3 shows the plot of the 1600 cm^{-1} Raman peak areas for leaves 3, 4, and 5 from KI110 and Native plants. One-way ANOVA was performed for rosmarinic acid abundance within leaves of different ages for epidermal cells in KI110 plants (Figure 3A), trichomes in KI110 plants (Figure 3B), epidermal cells in Native plants (Figure 3C), and trichomes in Native plants (Figure 3D). There were no

statistical differences in the 1600 cm^{-1} Raman peak areas for leaves with different ages within each plant and cell type, which indicates that any of the leaves (leaf 3, 4 or 5) may be measured without changing the results of the analysis.

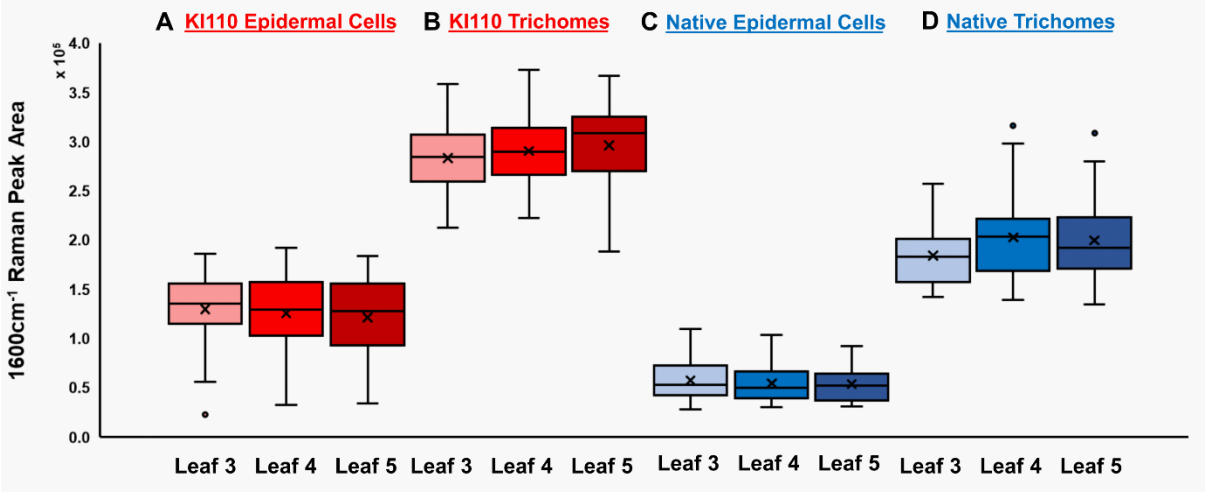

Figure 3. Box and whisker plot of the rosmarinic acid abundance determined from the 1600 cm^{-1} peak area within leaves of different ages in: (A) KI110 epidermal cells, (B) KI110 trichomes, (C) Native epidermal cells, and (D) Native trichomes. No statistically significant differences were measured with leaves of different ages; the *p*-values obtained from one-way ANOVA are shown in Supplemental Table 4.

3.4 Rosmarinic Acid Abundance Determined from Raman Spectra with Principal Component Analysis (PCA)

The aforementioned regression analysis required fitting the Raman peak area, which is suitable for some purposes but may not be suitable for rapid screening of a large number of plants. Thus, a chemometric model that required minimal spectral processing was developed to differentiate plants by rosmarinic acid abundance.

Principle component analysis (PCA) was pursued because it is a simple model that is useful for dimensionality reduction. As shown in Figure 4A, the percentage of variance in the Raman data that is explained in principal component (PC) 1, PC2, and PC3 is 77.89%, 9.42%, and 5.43%, respectively. The majority of the variance in PC1 is in the spectral region around 1600 cm^{-1} as shown in the PCA loadings in Figure 4B, suggesting that PC1 may be suitable for determining relative rosmarinic acid abundance. As expected, PC1 does order the spectra based on the rosmarinic acid abundance measured by the 1600 cm^{-1} peak area (i.e., the average PC1 for KI110 trichomes > PC1 for Native trichomes > PC1 for KI110 epidermal cells > PC1 for Native epidermal cells, Figure 4C). Of course there is considerable spread in the PC1 value for a given plant and cell type, which likely represents a spread in the rosmarinic acid abundance given the data in Figure 2B. However, it is not possible to confirm this as each data point in Figure 4C represents the Raman analysis of a few leaves, and there is not enough sample to perform an HPLC analysis on a few leaves. The important conclusion is that PC1 can be used as a rapid screen of relative rosmarinic acid abundance in spearmint plants.

PC2 (Figure 4C) and PC3 were also analyzed (Supplemental Figure 2). PC2 and PC3 partially separated the data based on Native and KI110 plant types. Thus the variance captured in these PCs is most likely representative of plant type and only to a lesser extent capturing variance based on rosmarinic acid abundance. This may be useful in some applications, but was not the goal of this work.

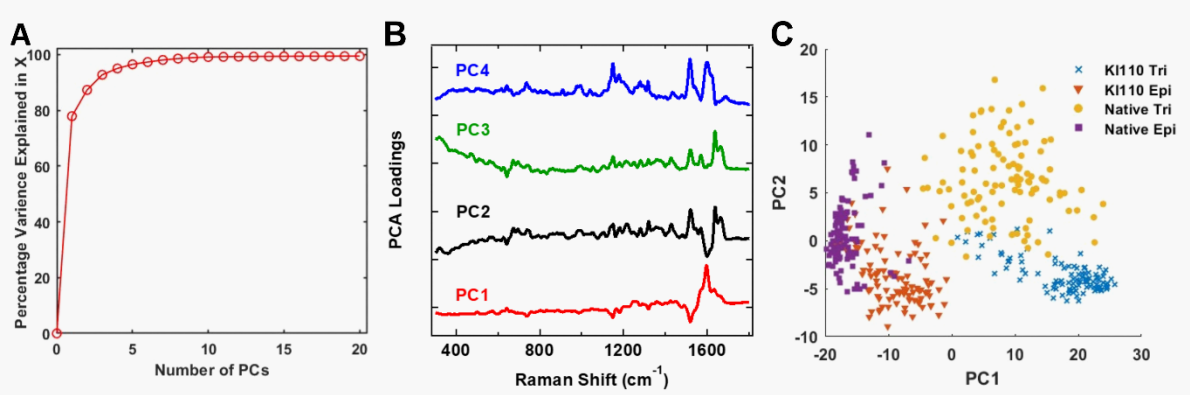

Figure 4. Results from PCA analysis. (A) Percentage of the variance explained in the principal components (PCs) of the Raman spectra. (B) PCA loadings of PC1-PC4. (C) Score plot of PC1 and PC2.

In summary, this work reports the correlation between Raman data collected from leaves and the rosmarinic acid abundance in spearmint plants. The linear regression model using the Raman peak area and PCA model show the most promise in supporting large scale screens for determining optimal harvest time or to advance plant breeding efforts. Discovering plant variants with high yields of a desired metabolite generally requires screening large plant populations.[43-45] Raman spectroscopy has the potential to be used as a low-cost and rapid preliminary screening method to discover useful variants that may benefit plant breeding efforts, although additional work in support of these efforts is required. These studies used a laboratory-grade Raman instrument so that the Raman spectra could be spatially mapped on the leaf. Raman spectroscopy is compatible with remoting sensing and handheld devices, so it may be possible to measure the rosmarinic acid abundance in the field without any sample preparation and without damaging the plants. A suggested future avenue to investigate is the use of a handheld Raman spectrometer for measurements in the field or greenhouse. A handheld Raman spectrometer typically has a

working distance of about 5-10 mm and a spot size in a range from several microns to several millimeters. The focus of this work is rosmarinic acid in spearmint, but models can also be envisioned for future rapid *in situ* Raman measurements to quantify the concentration of other metabolites in other plant types. Similarly, while the focus of this work is leaves, we have also shown that Raman spectroscopy can be used for the analysis of stems and roots.[46]

Acknowledgments

The authors thank Dr. Yumou Qiu from the statistics department at Iowa State University for his guidance with the statistical interpretations of data presented in this work.

Funding

The U.S. Department of Energy supported this research, Office of Science, Office of Biological and Environmental Research (BER) through Ames Laboratory. The Ames Laboratory is operated for the U.S. Department of Energy by Iowa State University under Contract No. De-AC02-07CH11358.

Conflict of Interest Statement

Kemin Industries, Inc., has a financial interest in using the developed methods.

Supporting information contains additional figures and tables for the main manuscript file.

References

1. Akula, R.; Ravishankar, G. A., Influence of abiotic stress signals on secondary metabolites in plants. *Plant signaling & behavior* **2011,** *6* (11), 1720-1731.
2. Mithöfer, A.; Schulze, B.; Boland, W., Biotic and heavy metal stress response in plants: evidence for common signals. *FEBS Letters* **2004,** *566* (1), 1-5.
3. Schulz, H.; Baranska, M., Identification and quantification of valuable plant substances by IR and Raman spectroscopy. *Vibrational Spectroscopy* **2007,** *43* (1), 13-25.
4. Qin, J.; Kim, M. S.; Chao, K.; Dhakal, S.; Cho, B.-K.; Lohumi, S.; Mo, C.; Peng, Y.; Huang, M., Advances in Raman spectroscopy and imaging techniques for quality and safety inspection of horticultural products. *Postharvest Biology and Technology* **2019,** *149*, 101-117.
5. Chandler, L.; Huang, B.; Mu, T. T. In *A smart handheld Raman spectrometer with cloud and AI deep learning algorithm for mixture analysis*, Next-Generation Spectroscopic Technologies XII, International Society for Optics and Photonics: 2019; p 1098308.
6. Pelletier, M., Quantitative analysis using Raman spectrometry. *Applied spectroscopy* **2003,** *57* (1), 20A-42A.
7. Gray, S. R.; Peretti, S. W.; Lamb, H. H., Real-time monitoring of high-gravity corn mash fermentation using in situ raman spectroscopy. *Biotechnology and bioengineering* **2013,** *110* (6), 1654-1662.
8. Iversen, J. A.; Ahring, B. K., Monitoring lignocellulosic bioethanol production processes using Raman spectroscopy. *Bioresource technology* **2014,** *172*, 112-120.
9. Shih, C.-J.; Lupoi, J. S.; Smith, E. A., Raman spectroscopy measurements of glucose and xylose in hydrolysate: role of corn stover pretreatment and enzyme composition. *Bioresource technology* **2011,** *102* (8), 5169-5176.
10. Shih, C.-J.; Smith, E. A., Determination of glucose and ethanol after enzymatic hydrolysis and fermentation of biomass using Raman spectroscopy. *Analytica chimica acta* **2009,** *653* (2), 200-206.
11. Ewanick, S.; Schmitt, E.; Gustafson, R.; Bura, R., Use of Raman spectroscopy for continuous monitoring and control of lignocellulosic biorefinery processes. *Pure and Applied Chemistry* **2014,** *86* (5), 867-879.
12. Ewanick, S. M.; Thompson, W. J.; Marquardt, B. J.; Bura, R., Real-time understanding of lignocellulosic bioethanol fermentation by Raman spectroscopy. *Biotechnology for biofuels* **2013,** *6* (1), 1-8.
13. Lawaetz, A. J.; Christensen, S. M.; Clausen, S. K.; Jørnsgaard, B.; Rasmussen, S. K.; Andersen, S. B.; Rinnan, Å., Fast, cross cultivar determination of total carotenoids in intact carrot tissue by Raman spectroscopy and partial least squares calibration. *Food Chemistry* **2016,** *204*, 7-13.
14. Anastasaki, E. G.; Kanakis, C. D.; Pappas, C.; Maggi, L.; Zalacain, A.; Carmona, M.; Alonso, G. L.; Polissiou, M. G., Quantification of crocetin esters in saffron (Crocus sativus L.) using Raman spectroscopy and chemometrics. *Journal of agricultural and food chemistry* **2010,** *58* (10), 6011-6017.
15. Lee, K.-M.; Herrman, T. J.; Yun, U., Application of Raman spectroscopy for qualitative and quantitative analysis of aflatoxins in ground maize samples. *Journal of Cereal Science* **2014,** *59* (1), 70-78.
16. Almeida, M. R.; Fidelis, C. H.; Barata, L. E.; Poppi, R. J., Classification of Amazonian rosewood essential oil by Raman spectroscopy and PLS-DA with reliability estimation. *Talanta* **2013,** *117*, 305-311.
17. Khodabakhshian, R.; Abbaspour-Fard, M. H., Pattern recognition-based Raman spectroscopy for non-destructive detection of pomegranates during maturity. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **2020,** *231*, 118127.

18. Pérez, M. R. V.; Mendoza, M. G. G.; Elías, M. G. R.; González, F. J.; Contreras, H. R. N.; Servín, C. C., Raman spectroscopy an option for the early detection of citrus Huanglongbing. *Applied Spectroscopy* **2016,** *70* (5), 829-839.
19. Zhang, X.; Qi, X.; Zou, M.; Liu, F., Rapid authentication of olive oil by Raman spectroscopy using principal component analysis. *Analytical letters* **2011,** *44* (12), 2209-2220.
20. Li, Y.-S.; Church, J. S., Raman spectroscopy in the analysis of food and pharmaceutical nanomaterials. *Journal of food and drug analysis* **2014,** *22* (1), 29-48.
21. Yang, D.; Ying, Y., Applications of Raman spectroscopy in agricultural products and food analysis: A review. *Applied Spectroscopy Reviews* **2011,** *46* (7), 539-560.
22. Luo, C.; Zou, L.; Sun, H.; Peng, J.; Gao, C.; Bao, L.; Ji, R.; Jin, Y.; Sun, S., A Review of the Anti-Inflammatory Effects of Rosmarinic Acid on Inflammatory Diseases. *Frontiers in Pharmacology* **2020,** *11*, 153.
23. Ly, T. N.; Shimoyamada, M.; Yamauchi, R., Isolation and characterization of rosmarinic acid oligomers in Celastrus hindsii Benth leaves and their antioxidative activity. *Journal of agricultural and food chemistry* **2006,** *54* (11), 3786-3793.
24. Kelm, M.; Nair, M.; Strasburg, G.; DeWitt, D., Antioxidant and cyclooxygenase inhibitory phenolic compounds from Ocimum sanctum Linn. *Phytomedicine* **2000,** *7* (1), 7-13.
25. Martins-Gomes, C.; Nunes, F. M.; Sampaio, A.; Souto, E. B.; Silva, A. M., Rosmarinic acid: Sources, bioactivities and health benefits. *Phytochemicals: Plant Sources and Potential Health Benefits; Ryan, I., Ed* **2019**, 109-146.
26. Lee, S. Y.; Xu, H.; Kim, Y. K.; Park, S. U., Rosmarinic acid production in hairy root cultures of Agastache rugosa Kuntze. *World Journal of Microbiology and Biotechnology* **2008,** *24* (7), 969-972.
27. Sik, B.; Kapcsándi, V.; Székelyhidi, R.; Hanczné, E. L.; Ajtony, Z., Recent advances in the analysis of rosmarinic acid from herbs in the Lamiaceae family. *Natural Product Communications* **2019,** *14* (7), 1934578X19864216.
28. Shekarchi, M.; Hajimehdipoor, H.; Saeidnia, S.; Gohari, A. R.; Hamedani, M. P., Comparative study of rosmarinic acid content in some plants of Labiatae family. *Pharmacognosy magazine* **2012,** *8* (29), 37.
29. Fletcher, R. S.; Slimmon, T.; Kott, L. S., Environmental factors affecting the accumulation of rosmarinic acid in spearmint (Mentha spicata L.) and peppermint (Mentha piperita L.). *The Open Agriculture Journal* **2010,** *4* (1).
30. Greaves, J. A.; Narasimhamoorthy, B.; Wildgen, S.; Barkley, R.; Ruden, S., Spearmint plant denominated KI-MsEM0110. Google Patents: 2017.
31. Narasimhamoorthy, B.; Zhao, L.; Liu, X.; Yang, W.; Greaves, J., Differences in the chemotype of two native spearmint clonal lines selected for rosmarinic acid accumulation in comparison to commercially grown native spearmint. *Industrial Crops and Products* **2015,** *63*, 87-91.
32. Schulze, G.; Jirasek, A.; Marcia, M.; Lim, A.; Turner, R. F.; Blades, M. W., Investigation of selected baseline removal techniques as candidates for automated implementation. *Applied spectroscopy* **2005,** *59* (5), 545-574.
33. Mariappan, G.; Sundaraganesan, N.; Manoharan, S., Experimental and theoretical spectroscopic studies of anticancer drug rosmarinic acid using HF and density functional theory. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **2012,** *97*, 340-351.
34. Afseth, N. K.; Segtnan, V. H.; Wold, J. P., Raman spectra of biological samples: A study of preprocessing methods. *Applied spectroscopy* **2006,** *60* (12), 1358-1367.
35. Barnes, R.; Dhanoa, M. S.; Lister, S. J., Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied spectroscopy* **1989,** *43* (5), 772-777.
36. Lupoi, J. S.; Gjersing, E.; Davis, M. F., Evaluating lignocellulosic biomass, its derivatives, and downstream products with Raman spectroscopy. *Frontiers in bioengineering and biotechnology* **2015,** *3*, 50.

37. Kokkini, S.; Karousou, R.; Lanaras, T., Essential oils of spearmint (Carvone-rich) plants from the island of Crete (Greece). *Biochemical Systematics and Ecology* **1995,** *23* (4), 425-430.
38. Schulz, H.; Özkan, G.; Baranska, M.; Krüger, H.; Özcan, M., Characterisation of essential oil plants from Turkey by IR and Raman spectroscopy. *Vibrational Spectroscopy* **2005,** *39* (2), 249-256.
39. Chauhan, R.; Kaul, M.; Shahi, A.; Kumar, A.; Ram, G.; Tawa, A., Chemical composition of essential oils in Mentha spicata L. accession [IIIM (J) 26] from North-West Himalayan region, India. *Industrial crops and products* **2009,** *29* (2-3), 654-656.
40. Wu, Z.; Tan, B.; Liu, Y.; Dunn, J.; Martorell Guerola, P.; Tortajada, M.; Cao, Z.; Ji, P., Chemical Composition and Antioxidant Properties of Essential Oils from Peppermint, Native Spearmint and Scotch Spearmint. *Molecules* **2019,** *24* (15), 2825.
41. Heraud, P.; Beardall, J.; McNaughton, D.; Wood, B. R., In vivo prediction of the nutrient status of individual microalgal cells using Raman microspectroscopy. *FEMS microbiology letters* **2007,** *275* (1), 24-30.
42. Wiley, J. H.; Atalla, R. H., Band assignments in the Raman spectra of celluloses. *Carbohydrate Research* **1987,** *160*, 113-129.
43. Fletcher, R. S.; McAuley, C.; Kott, L. S. In *Novel Mentha spicata clones with enhanced rosmarinic acid and antioxidant activity*, III WOCMAP Congress on Medicinal and Aromatic Plants-Volume 6: Traditional Medicine and Nutraceuticals 680, 2003; pp 31-36.
44. Kott, L.; Fletcher, R., Production of Rosmarinic Acid from Spearmint and uses Thereof. Google Patents: 2010.
45. Fletcher, R. S.; Slimmon, T.; McAuley, C. Y.; Kott, L. S., Heat stress reduces the accumulation of rosmarinic acid and the total antioxidant capacity in spearmint (Mentha spicata L). *Journal of the Science of Food and Agriculture* **2005,** *85* (14), 2429-2436.
46. Burkhow, S. J.; Stephens, N. M.; Mei, Y.; Dueñas, M. E.; Freppon, D. J.; Ding, G.; Smith, S. C.; Lee, Y.-J.; Nikolau, B. J.; Whitham, S. A., Characterizing virus-induced gene silencing at the cellular level with in situ multimodal imaging. *Plant methods* **2018,** *14* (1), 1-12.

For Table of Contents Only