Analytical and Bioanalytical Chemistry
https://doi.org/10.1007/s00216-020-02706-5

RESEARCH PAPER

Non-invasive identification of potato varieties and prediction of the origin of tuber cultivation using spatially offset Raman spectroscopy

Rohini Morey[1] · Alexei Ermolenkov[1] · Willam Z. Payne[1] · Douglas C. Scheuring[2] · Jeffrey W. Koym[3] · M. Isabel Vales[2] · Dmitry Kurouski[1,4]

Received: 30 March 2020 / Revised: 5 May 2020 / Accepted: 13 May 2020

Abstract

High starch content, simplicity of cultivation, and high productivity make potatoes (*Solanum tuberosum*) a staple in the diet of people around the world. On average, potatoes are composed of 83% water and 12% carbohydrates, and the remaining 4% includes proteins, vitamins, and other trace elements. These proportions vary depending on the type of potato and location where they were cultivated. At the same time, the chemical composition determines the nutritional value of potato tubers and can be proved using various wet chemistry and spectroscopic methods. For instance, gravity measurements, as well as several different colorimetric assays, can be used to investigate the starch content. However, these approaches are indirect, often destructive, and time- and labor-consuming. This study reports on the use of Raman spectroscopy (RS) for completely non-invasive and non-destructive assessment of nutrient content of potato tubers. We also show that RS can be used to identify nine different potato varieties, as well as determine the origin of their cultivation. The portable nature of Raman-based identification of potato offers the possibility to perform such analysis directly upon potato harvesting to enable quick quality evaluation.

Keywords Potato varieties · Raman spectroscopy · Phenotyping · Nutrient content · SORS · Identification

Introduction

Potatoes (*Solanum tuberosum*) were originally domesticated in Southern Peru, Northwestern Bolivia, around 5000 B.C. After the Columbian exchange of agricultural goods with the New World, potatoes spread across Europe and Asia. Nowadays, potatoes are grown in more than 160 countries worldwide [1]. Substantial differences in climate and soil profiles required the development of new potato varieties that would produce high yields in different sites around the world. As a result of centuries of crossbreeding, there are thousands of varieties of potatoes grown on six different continents.

Currently, potato variety identification is carried by horticultural experts and potato breeders based on morphological traits (plants and tubers) and molecular characterization (mainly DNA-based) [2, 3]. This expertise requires substantial academic training and professional experience. DNA-based fingerprinting is a good complementary (less subjective, and not affected by the environment) alternative to the morphological identification of potato. However, DNA-based fingerprinting is time and labor-consuming and not portable. These limitations substantially limit the use of DNA-fingerprinting for routine variety identification in agriculture [4].

In addition to identification, it is highly important to determine nutrient content of potato. There are several wet laboratory assays that can be used to assess the nutrient composition of tubers [5]. For instance, starch content in potato can be determined using colorimetric/fluorometric assays [6]. The colorimetric approach is based on enzymatic conversion of

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00216-020-02706-5) contains supplementary material, which is available to authorized users.

✉ Dmitry Kurouski
dkurouski@tamu.edu

[1] Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX 77843, USA

[2] Department of Horticultural Sciences, Texas A&M University, College Station, TX 77843, USA

[3] Texas A&M AgriLife Research and Extension Center, Lubbock, TX 79403, USA

[4] The Institute for Quantum Science and Engineering, Texas A&M University, College Station, TX 77843, USA

Published online: 25 May 2020

starch into a colored reaction product. The change in color is then measured using absorption spectroscopy. The Dumas combustion method is typically used to probe protein content via analysis of the total nitrogen amount in the sample [7]. Although broadly used, these approaches are destructive, time, and labor-consuming. It has been suggested that protein and starch content of potato can be determined using Near-infrared (NIR) spectroscopy [8–12]. However, the use of NIR is limited to only freeze-dry samples, as the accuracy of this approach drastically decreases with an increase in water content in the sample. Therefore, NIR did not find broad applicability for the analysis of nutrient content of potatoes.

A useful and efficient alternative is presented in Raman Spectroscopy (RS) [4]. RS is an analytical technique that is based on inelastic light scattering of molecules that are being excited to higher vibrational or rotational states [13]. Once collected, the inelastically scattered photons provide information about the chemical structure of the sample. We have previously shown that RS could be used for confirmatory identification of six different varieties of corn as well as for determination of their nutrient content [14]. One could envision that this spectroscopic approach opened a new paradigm in the evaluation of the economic value of agricultural products. The question to ask is whether RS can be used to identify different potato varieties, as well as to access the nutrient content of potato tubers. The main components of the average potato tuber are water (83%), carbohydrates (~ 12%), and proteins (4%) and other trace nutrients (~ 1%) [15]. These proportions vary depending on the type of potato and location where they were cultivated. In the current study, we used a hand-held Raman spectrometer to investigate the possibility of Raman-based identification of nine different potato varieties. We have also explored the possibility to track the geographic origin of potato cultivation based on spectroscopic signatures of potato tubers.

Materials and methods

Potato varieties and locations

All potato varieties (Table 1) were planted and harvested in the State of Texas at two different locations: Springlake (34° 8′ 6.97″ N, 102°21′ 51.18″ W) and Dalhart (35° 58′ 15.31″ N, 102° 44′ 36.33″ W), which are both in Northwest Texas. The potatoes were planted in a randomized complete block design with four replications, tuber seed pieces at a 15.2-cm depth in the field. Each plot consisted of two rows with a total number of 28 plants (Springlake) and 24 plants (Dalhart) (Electronic Supplementary Material (ESM) Table S1). The potato seed was protected (seed treatment) before planting. Fertilizers, insecticides, fungicides, and herbicides were applied to ensure their growth (details included in the ESM). They were irrigated using a center pivot system and then were harvested using a two-row drag digger and bagged by hand. The potato growing locations were subject to different environmental conditions (Table 1). At harvest time, a sample of ten tubers per rep per location per clone was collected. Weight per tuber ranged between 113.4–170.1 g (this corresponds to a tuber grading class within marketable tubers). The tubers were maintained at room temperature (22 °C) until it was time to analyze them.

Raman spectroscopy

A hand-held Agilent Resolve SORS instrument was used to obtain spectra from the potato samples. This spectrometer has a spectral resolution of 15 cm^{-1} and an 830-nm laser excitation. The instrument collects three types of spectra: “surface,” “offset,” and “spatially offset Raman spectra (SORS)” [16]. The “surface” spectra are collected from the surface of the sample. We avoided using surface spectra for our study because of possible soil or organic contaminants that could be present on the surface of analyzed potato tubers. One can imagine that such contaminants could cause drastic spectroscopic changes and, consequently, abstract analyses of the nutrient content of tubers. The “offset” is the spectra acquired from the inner part of the sample, while “SORS” denotes the difference between the surface and the offset that is performed by the automated subtraction of “surface” from “offset” spectra. If the chemical structure of the surface and inner part of the specimen is different, such spectral subtraction becomes extremely practical to enable clear visualization of a spectroscopic signature of the inner part. This approach is used for the detection of drugs and explosives in non-transparent containers and bags [17–19]. However, if the chemical structure of the surface and inner part of the sample is similar, like in the case of potato, the instrument does not perform an accurate spectral subtraction to generate consistent SORS results. Therefore, in our study, we used the ‘offset’ rather than SORS data. For each offset scan, an offset position of 3.5 to 4 mm (depending on the variety) was used, along with a 2-s integration time with 30 accumulations. Spectra were taken from the intact potato surface without peeling or cutting of the tuber. Every tuber was scanned a couple of times in spatially non-overlapping locations. There were 18 sets of data, each categorized by variety and location, and 40 to 60 spectra were acquired from each data set. The spectra were then averaged and processed using multivariate statistical analysis.

It should be noted that analyzed potato tubers had substantial differences in coloration (some tubers had dark color and some light yellow color). One can imagine that dark color tubers will absorb more and, consequently, scatter less light compared to the yellow or pale color tubers. Since RS is based on inelastic light scattering, dark color tubers should produce

less intense spectra (under the same experimental conditions) compared with the light color tubers (ESM Fig. S1). Therefore, observed variations in spectral intensities likely originate from different light absorption and scattering properties of these tubers. This evidence suggests that raw spectra cannot be directly used for analysis of nutrient content of potatoes (discussed below). This problem can be solved by spectral normalization. Therefore, we normalized all reported spectra in this manuscript on the intensity of 1460 cm^{-1}. This vibrational band can be assigned to CH and CH_2 vibrations that are present in nearly all classes in biological molecules, which makes such normalization the least biased to any of the analyzed classes of nutrients.

Multivariate statistical analysis and analysis of variance

The PLS_Toolbox software was used to perform partial least-squares discriminant analysis (PLS-DA) for all collected offset spectra. The spectra were pre-processed by taking the 2nd derivative of all intensity values (3rd polynomial order and a filter length of 51) and then centered on the mean. A true positive rate (TPR) was reported for each category based on the accuracy rate of the predictions. This process was used to determine the difference between the varieties and between the locations.

We also performed analysis of variance (ANOVA) in MATLAB to compare the content of starch, phenylpropanoids, and carotenoids between different potato varieties. Reported ANOVAs compared the intensity of peaks at 479 cm^{-1} and 1125 cm^{-1} to analyze the starch, 1600 and 1630 cm^{-1} phenylpropanoid, and the peak at 1527 cm^{-1} to access the carotenoid contents. Lastly, ANOVA on 1660 cm^{-1} was used to determine protein content in the analyzed potato tubers.

Starch content determination

Homogenous starch gel samples were made by mixing various amounts of potato starch with 50 mL distilled water and .25 mL dimethyl sulfoxide. The starch was measured to 4.5 g for the 9% sample, 6.0 g for the 12% solution, and so on. The solutions were then heated in an Erlenmeyer flask with a stir bar. When the proper temperature was reached, the solution turned to gel and was removed from heat. The gel was cooled overnight and scanned the next day using RS. The intensity of the peak at 479 cm^{-1} was measured and a calibration curve was made from the calculated averages and standard deviations of the data (ESM Fig. S2).

Results and discussions

We collected more than 400 spectra from tubers of nine different potato varieties. Six out of nine varieties are mainly used for the fresh potatoes, which tend to be sold directly to the consumers (ESM Table S1). They include Red LaSoda (RLS), Yukon Gold (YG), Russet Norkotah (RN), Sierra Rose/ATTX961014-1R/Y (SR), COTX04050s-1P/P (P/P), and PORTX03PG25-2R/R (R/R). The first three are commercial, public reference potato varieties with red, yellow, and russet skin, respectively, while the last three (also COTX09022-3RuRE/Y) were developed by the Texas A&M university Potato Breeding Program and have different combinations of skin and flesh color. The latter four of the nine varieties are processing varieties including Russet Burbank (RB) (used mainly to make French fries), Atlantic (AT) (used to make potato chips) and COTX09022-3RuRE/Y (Ru/Y) (dual purpose, fresh or processing), ESM Table S1.

Raman spectra collected from tubers exhibited vibrational bands originating from carbohydrates, carotenoids, phenylpropanoids, and proteins (Fig. 1, Table 2, ESM Fig.

Table 1 Names, abbreviations, and tuber descriptions of the potato varieties used in our study. The potato trials were planted at two locations in Texas (Springlake and Dalhart) in 2019

Clone name	Code	Source	Tuber skin color	Tuber flesh color	Tuber shape	Market class	Market subclass
Red LaSoda	RLS	Reference commercial	Red	White	Round	Fresh	
Yukon Gold	YG	Reference commercial	Yellow	Yellow	Round	Fresh	
Sierra Rose™ (ATTX961014-1R/Y)	SR	TAMU—commercial	Red	Yellow	oblong	Fresh	
COTX04050s-1P/P	P/P	TAMU—experimental	Purple	Purple	Round	Fresh	
PORTX03PG25-2R/R	R/R	TAMU—experimental	Red	Red	Fingerling	Fresh	
Russet Burbank	RB	Reference commercial	Russet	White	Oblong to long	processing	French fries
Russet Norkotah	RN	Reference commercial	Russet	White	Oblong to long	Fresh	
Atlantic	AT	Reference commercial	White	White	Round	Processing	Chipper
COTX09022-3RuRE/Y	Ru /Y	TAMU—commercial	Russet with red eyes	Yellow	Round to oblong	Dual purpose	

S1). We found that spectroscopic signatures of SR and YG appeared to be very similar, whereas Ru/Y exhibited lower intensities at 1054, 1085, and 1126 cm^{-1}. We have also observed that intensities of 1261 and 1340 cm^{-1} were lower in the spectrum of Ru/Y comparing with the intensities of these bands in the spectrum of SR and YG. Much stronger spectral differences have been observed between the spectra collected from P/P, R/R, and RLS. Next, we found that the spectra of AT exhibited much more intense carbohydrate bands (479, 865, 940, 1054, 1085, 1126, 1383–1398 cm^{-1}) when compared to the spectra of RB and RN. Lastly, we found that the intensities of the phenylpropanoid bands gradually decreased in the order of RN, RB, and AT. These spectral changes in intensities suggest variations of nutrient content in the tubers. Summarizing spectral changes observed for nine different potato varieties, we can conclude that we found a gradual decrease in the intensities of carbohydrate bands (479, 865, 940, 1054, 1085, and 1126 cm^{-1}) from SR and YG to RLS and Ru/Y to AT, P/P, R/R and RB to RN. We also found that spectra of P/P and R/R had a substantially large intensity of carotenoids (1527 cm^{-1}), relative to the intensity of this band in the spectra collected from other potato varieties. Intensity of phenylpropanoid bands (1600–1630 cm^{-1}) drastically decreased from the spectra of R/R to P/P and then to RLS. Next, we observed a graduate decrease in the intensity of phenylpropanoid bands from RLS to SR, RN, YG, RB, Ru/Y, and AT.

ANOVA was used to compare relative intensities of the described above vibrational bands, Fig. 2. Results demonstrate that the intensity of 1126 cm^{-1} band in the spectra of

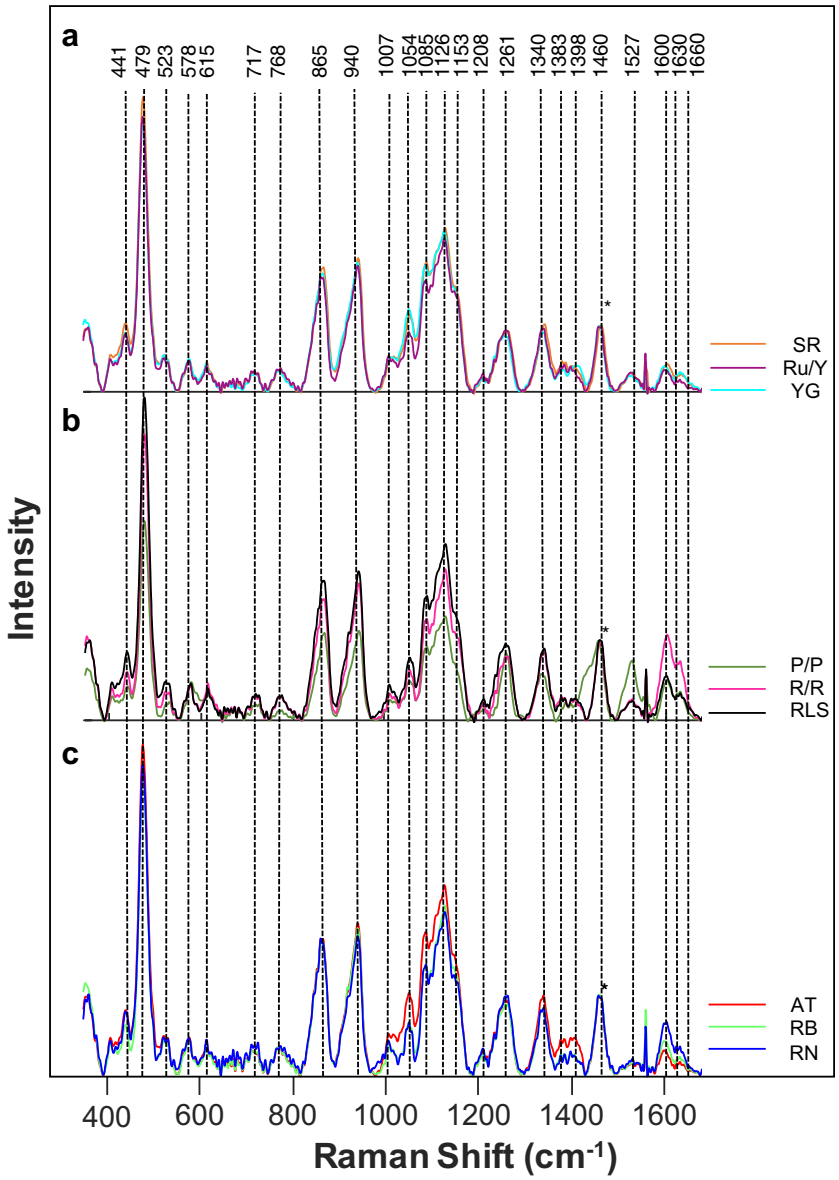

Fig. 1 Raman spectra of nine different potato varieties separated into three groups (**a**–**c**) for clarity of visualization. Asterisk denotes a 1460-cm^{-1} peak was used to normalize spectra

SR and YG are stronger when compared to the intensity of other potato varieties. This band was also found to have the lowest intensity in the spectrum of RN. This suggests that SR and YG have the highest starch content, whereas RN has the lowest starch content.

ANOVA of carotenoids content revealed the presence of two different groups of potato with respect to its carotenoid content. Specifically, P/P exhibited significantly larger carotenoid content relative to all other potato varieties, whereas carotenoid content in all other potato varieties is nearly identical. It is important to note that two out of eight analyzed potato varieties have colored flesh, whereas the other six varieties exhibit similar white flesh color. At the same time, only one of the colored potatoes (P/P) exhibited distinctly different carotenoid content. This finding indicates that carotenoids are present in this purple flesh potato and likely contribute to intensity the purple color, The purple color of P/P and red color of R/R is caused by presence of anthocyanins, molecules with distinctly different chemical structure. These results clearly demonstrate that Raman-based analysis of nutrient content cannot be misled by sample coloration. On the opposite, RS can be used to reveal the underlying cause of coloration and distinguish between different chromophores of potato.

Intensity of the peak at 1600 cm^{-1} can be used to estimate phenylpropanoid content in the potatoes. It can be noticed that R/R and P/P have a considerably higher intensity of 1600 cm^{-1} comparing with the rest of potato varieties. This suggests the highest content of phenylpropanoids in these two potato varieties. It should be noted that R/R and P/P clones are rich in anthocyanins (phenylpropanoids). The purple color is mainly due to delphinidin and the red color to pelargonin. AT and Ru/Y exhibit the lowest phenylpropanoid content. RB, RN, YG, SR, and RLS have slightly larger phenylpropanoid content relative to AT and Ru/Y and lower than P/P. RLS and SR are not statistically different from each other because of a small overlap of confidence intervals between the intensity of 1600 cm^{-1} band that has been observed between Raman spectra collected from those two potato varieties. One can conclude that from a perspective of phenylpropanoid content, all analyzed potato varieties fell into three statistically different classes: R/R with the highest phenylpropanoid content, P/P

Table 2 Assignment of chemical compounds based on the vibrational mode corresponding to each Raman shift wave value

Band (cm^{-1})	Vibrational mode	Assignment
441	Skeletal modes of pyranose ring	Carbohydrates [20, 21]
479	C–C–O and C–C–C deformations; Related to glycosidic ring skeletal deformations δ(C–C–C) + τ(C–O) Scissoring of C–C–C and out-of-plane bending of C–O	Carbohydrates [20]
523	δ(C–C–O) + τ(C–O) of carbohydrates	Carbohydrates [20]
578	ν(C–O) + ν(C–C) + δ(C–O–H)	Cellulose, phenylpropanoids [22]
615	δ(C–C–O) of carbohydrates	Carbohydrates [20]
717	δ(C–C–O) related to glycosidic ring skeletal deformations	Carbohydrates [20]
768	δ(C–C–O)	Carbohydrates [20]
865	δ(C–C–H) + δ(C–O–C) glycosidic bond; anomeric region	Carbohydrates [20]
940	Skeletal modes; δ(C–O–C) + δ(C–O–H) + ν(C–O) α-1,4 glycosidic linkages	Carbohydrates [23]
1007	In-plane CH3 rocking + C–C	Carotenoids [24]
1016	C–OH	Carbohydrates [25, 26]
1054	ν(C–O) + ν(C–C) + δ(C–O–H)	Carbohydrates [20]
1085	ν(C–O) + ν(C–C) + δ(C–O–H)	Carbohydrates [20]
1126	ν(C–O) + ν(C-C) + δ(C–O–H)	Carbohydrates [20]
1153	ν(C–O–C), ν(C–C) in glycosidic linkage, asymmetric ring breathing	Carbohydrates [27]
1208	aromatic ring modes of phenylalanine and tyrosine; symmetric O–CH3 wag + C–O–H bending	Proteins [28], Phenylpropanoids [29]
1261	δ(C–C–H) + δ(O–C–H) + δ(C–O–H)	Carbohydrates [20, 30]
1340	ν(C–O); δ(C–O–H)	Carbohydrates [20]
1383	δ(C–O–H)—coupling of the CCH and COH deformation modes	Carbohydrates [20]
1398	δ(C–C–H)	Carbohydrates [20]
1460	δ(CH) + δ(CH2) + δ(C–O–H) CH, CH2, and COH deformations.	Aliphatic [20]
1527	–C=C–	Carotenoids [31]
1600	ν(C–C) aromatic ring + σ(CH)	Phenylpropanoids [32, 33]
1630	C=C–C(ring)	Phenylpropanoids [34]
1660	amide I (C=O)	Proteins [13]

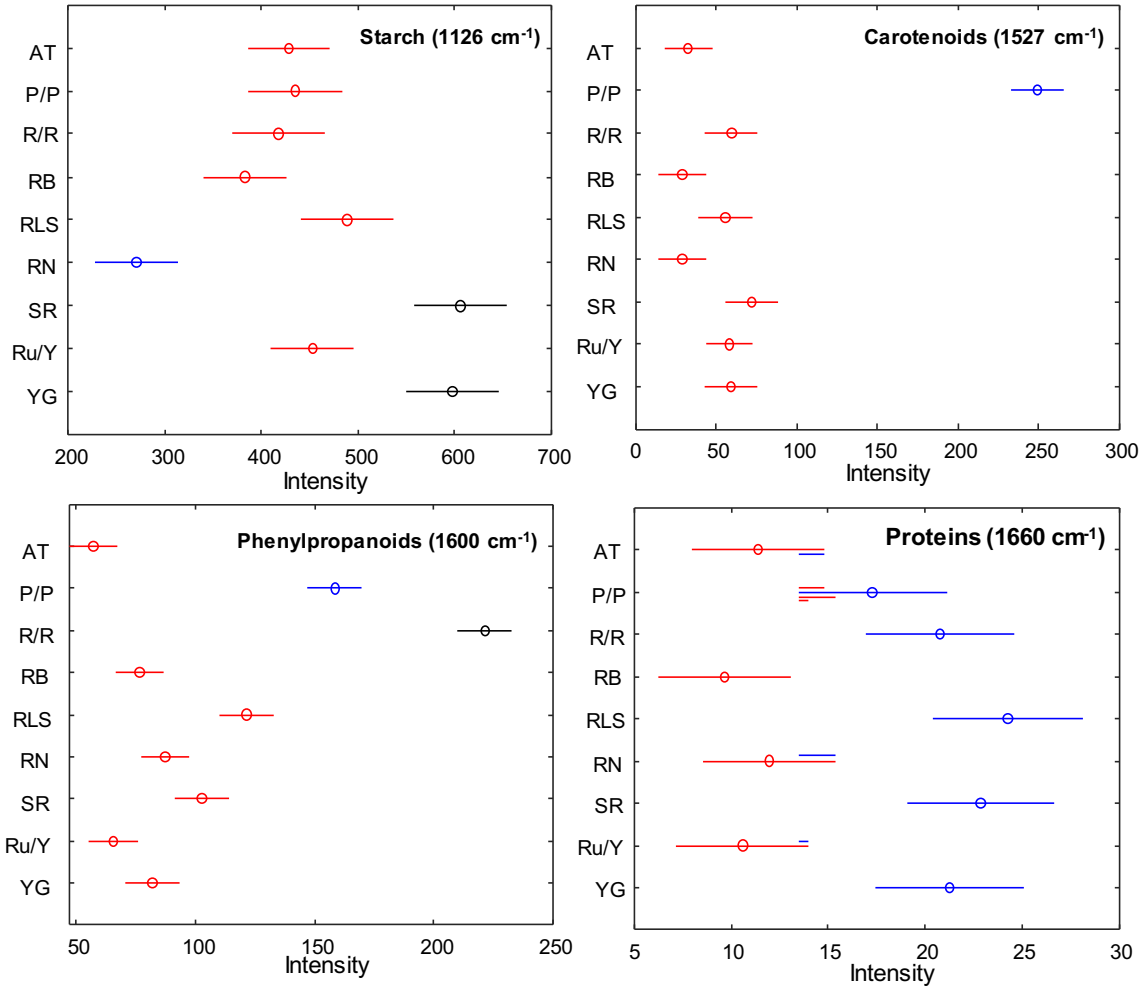

Fig. 2 Means (circles) and confidence intervals for the intensities of the potato spectra at 1126 cm^{-1} (starch), 1527 cm^{-1} (carotenoids), 1600 cm^{-1} (phenylpropanoids), and 1660 cm^{-1} (proteins). ANOVA of starch revealed 3 groups of potato varieties (blue, red, and black) with significantly different starch contents. ANOVA of carotenoids revealed 2 groups of potato varieties (blue and red) with significantly different carotenoid contents. ANOVA of phenylpropanoids revealed 3 groups of potato varieties (blue, red, and black) with significantly different phenylpropanoid contents. ANOVA of proteins revealed 2 groups of potato varieties (blue and red) with significantly different protein contents. Multiple colors indicate a member of a group that has overlap between two separate groups

with medium and AT, RB, RLS, RN, SR, Ru/Y and YG with low phenylpropanoid content.

Finally, the intensity of a vibrational band at 1660 cm^{-1} was used to determine the protein content in potato tubers.

Table 3 The average specific gravities and concentration of solids of each variety from both locations

Variety	Specific gravity	% solids
AT	1.065	14.4
P/P	1.064	13.8
R/R	1.059	13.0
RB	1.061	13.4
RLS	1.056	12.6
RN	1.059	13.1
SR	1.066	14.3
Ru/Y	1.074	15.7
YG	1.070	15.0

Table 4 The confusion table and percent accuracy for each variety

	N	TPR %	AT	P/P	R/R	RB	RLS	RN	SR	Ru/Y	YG
AT	60	65.0	39	0	0	10	2	1	2	6	0
P/P	39	100.0	0	39	0	0	0	1	0	0	0
R/R	40	100.0	0	0	40	1	2	0	0	0	0
RB	60	70.0	6	0	0	42	1	3	0	5	0
RLS	40	70.0	3	0	0	4	28	1	7	1	2
RN	58	84.5	2	0	0	0	0	49	0	9	0
SR	40	60.0	0	0	0	0	3	0	24	1	4
Ru/Y	60	63.3	9	0	0	3	1	3	2	38	0
YG	40	85.0	1	0	0	0	3	0	5	0	34

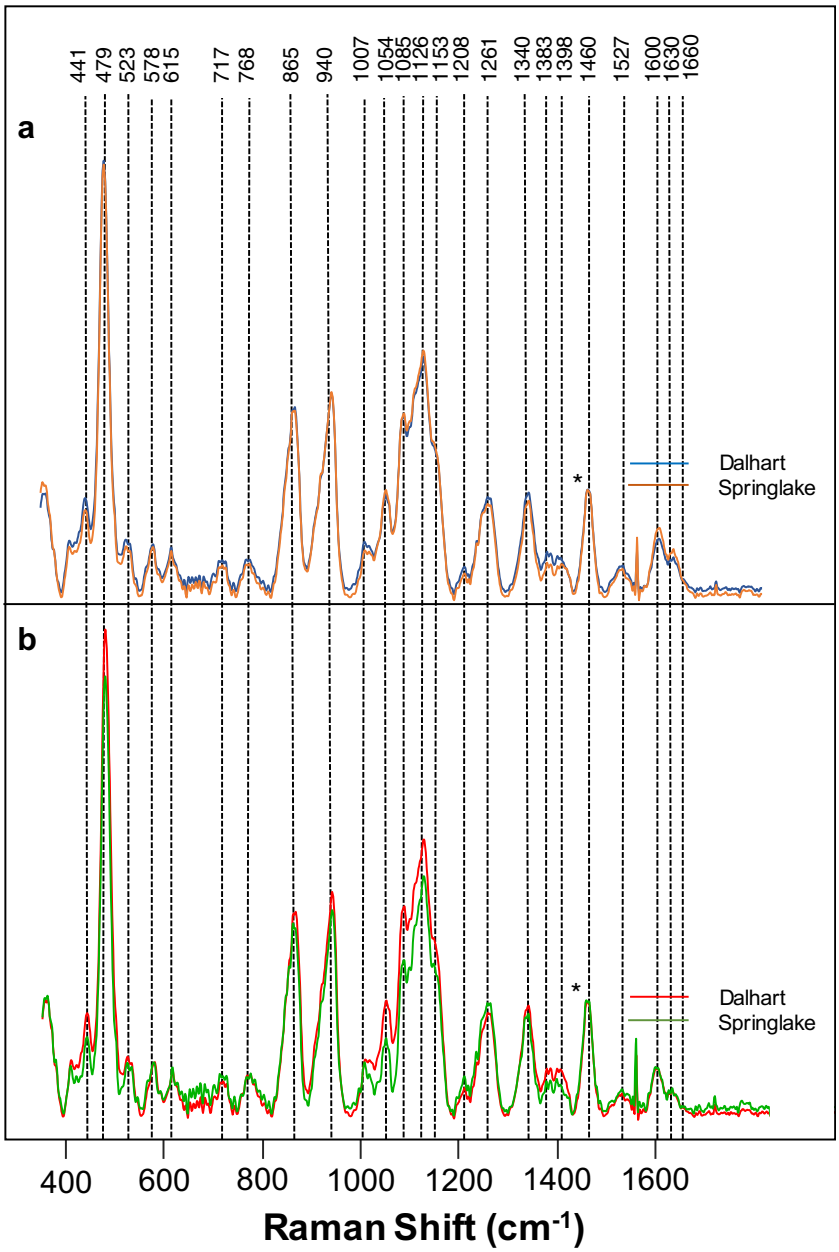

Fig. 3 Raman spectra of the fresh market (**a**) and chipper (**b**) potatoes grown in Springlake and Dalhart. TOC

ANOVA results showed that R/R, RLS, SR, and YG had a significantly higher content of proteins relative to RB potato variety. Concentration of protein in AT, P/P, RN, and Ru/Y was not significantly higher relative to RB and not significantly lower comparing with R/R, RLS, SR, and YG. Nevertheless, visual analysis of ANOVA results suggested that all nine potato varieties could be divided into two classes based on their protein content. AT, RB, RN, and Ru/Y could be assigned to class 1 (low protein content), whereas P/P, R/R, RLS, SR, and YG belong to class 2 (high protein content).

Table 5 Confusion table for the fresh market potatoes and the percent accuracies for each location

	N	TPR %	Springlake	Dalhart
Springlake	160	89.4	143	25
Dalhart	159	84.3	17	134

Table 6 Confusion table for the chipper potatoes and the percent accuracies for each location

	N	TPR %	Springlake	Dalhart
Springlake	238	89.9	214	41
Dalhart	155	90.9	24	141

Table 7 Percent accuracy values in percentages for each variety when they are compared separately for each location

Variety	Springlake	Dalhart
SR	100	81.6
YG	89.5	87.2
R/R	94.9	82.5
RLS	97.4	81.1

A question to ask is whether the reported above analysis of nutrient contact is truly unbiased if only one band from all different bands exhibited by a certain class of molecules is analyzed. For instance, we have analyzed the change in intensity of 1126 cm^{-1} band to evaluate differences in carbohydrate content, while these molecules also have bands at 479, 717, 768, 865, 940, 1054, and 1153 cm^{-1}. In addition to 1600 cm^{-1}, phenylpropanoids exhibit a band at 1630 cm^{-1}, whereas carotenoids also have a band at 1000 cm^{-1}. To investigate this, ANOVAs were performed on 479 and 1630 cm^{-1}. Results demonstrate nearly identical distribution of intensities of these bands for different potato varieties, as has been observed for 1126 and 1600 cm^{-1}, respectively, ESM Fig. S3. It should be noted that the use of 1126 cm^{-1} band in ANOVA allows for better differentiation between RLS and YG relative to 479 cm^{-1}. At the same time, the use of 1600 cm^{-1} enabled better differentiation between RLS and SR comparing with the separation achieved on ANOVA of 1630 cm^{-1}. These results suggest that 1126 cm^{-1} should be used for ANOVA of carbohydrate content of potatoes, whereas 1600 cm^{-1} is better suitable for the analysis of phenylpropanoids. It should be noted that a band at 1000 cm^{-1} cannot be used as the alternative to 1527 cm^{-1} for the assessment of carotenoid content because this band also present in the Raman spectra of proteins. These findings demonstrate that ANOVA of vibrational bands can be used for accurate and reliable assessment of nutrient content of potato varieties.

Potatoes mostly consist of starch and it has to be verified of whether the results shown in the ANOVAs are accurate. Note that approximately 83% of the average potato contains water, and the rest is mostly starch. Keeping this in mind, we can conclude that the specific gravity of the potato corresponds with the concentration of starch in water for each potato. The average specific gravity for each variety from both locations has been shown in Table 3. We can also look at the concentration of solids in the potato to analyze starch content. This information has been added ("% solids" category) in Table 3.

Analysis of specific gravities shows high similarity of prediction of starch content to the results obtained by RS (Fig. 2). We also demonstrated that the intensity of a 479-cm^{-1} band directly correlates with the starch content of the sample. For this, four samples with 9, 12, 15, and 18% of starch were analyzed by RS, ESM Fig. S2. Our results showed that the intensity of 479/1126 cm^{-1} linearly increased with an increase in the starch content of the sample. These results demonstrate that RS can be used for highly accurate determination of the starch content of the sample.

One may wonder whether such a clear difference in spectra collected from different potato varieties can be used to identify them and distinguish different potato varieties based on their spectra. We built PLS-DA model that showed that some potato varieties (P/P and R/R) can be identified with very high accuracy (near 100%), whereas the accuracy of potato variety prediction ranged from 60% (SR) to 85% (YG), Table 4, ESM Fig. S4. Specifically, PLS-DA showed that all 39 spectra of P/P and all 40 spectra of R/R were predicted correctly, which resulted in 100% accuracy of prediction of these two potato varieties. At the same time, 1 out of 40 YG spectra was predicted as AT, 3 as RLS, and 5 as SR, which resulted in a total of 85% accuracy prediction of YG based on the developed PLS-DA model. Thus, the higher number of spectra that were miss-predicted by the PLS-DA model for each of the potato varieties, the lower prediction score (TPR) was achieved for that class of spectra. These results demonstrate that RS can be used for automated identification of potato. Such potato identification can be used to sort potato upon harvesting. It can also be used to digitize potato warehouses and storage facilities, enabling automated identification of potato upon import and export of potato tubers.

We also questioned whether RS could be used to identify location where potato was grown. To investigate this, we compared spectra collected from the two different potato types (market and chipper) grown in two different locations: Springlake and Dalhart, both located in Northwest Texas (Fig. 3, Tables 5 and 6). Potato grown in these two locations experienced unique environmental conditions originating from differences in weather and soil. One can envision that such environmental differences could result in small changes in the nutrient composition of potato. Our results show that PLS-DA is able to distinguish potato grown in Springlake and Dalhart with 84.3% to 90.9% accuracy, ESM Fig. S5. These findings demonstrate that RS can be used to predict the geographical location of potato farming.

Next, we compared the accuracy of the prediction of four individual potato varieties grown in Springlake and Dalhart, Table 7. Our results show that the accuracy of prediction ranged from 81.1% (RLS, Dalhart) to 100% (SR, Springlake), ESM Fig. S6. Thus, the accuracy of prediction varies depending on the potato variety and location. For instance, differentiation of four different potato varieties has been higher for Springlake rather than for Dalhart. Summarizing, we found that PLS-DA can be used for highly accurate prediction whether individual potato varieties (SR, YG, R/R, and RLS) were grown in Springlake or Dalhart.

In conclusion, we demonstrated that RS could be used to accurately analyze potato nutrient composition, for potato variety identification, and determine the location of their origin.

The non-invasive and non-destructive nature of this analysis, together with the portability of the spectrometer [35–40], opens up a broad spectrum of opportunities for the use of RS in potato farming.

Funding information This study was financially supported by the AgriLife Research of Texas A&M and the Governor's University Research Initiative (GURI) grant program of Texas A&M University, GURI Grant Agreement No. 12-2016, M1700437.

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing interests.

References

1. Hijmans R. Global distribution of the potato crop. J Am J Potato Res. 2001;78:403–12.
2. Tillault AS, Yevtushenko DP. Simple sequence repeat analysis of new potato varieties developed in Alberta, Canada. Plant Direct. 2019;3:e00140. https://doi.org/10.1002/pld3.140.
3. Potato Genome Sequencing C, Xu X, Pan S, Cheng S, Zhang B, Mu D, et al. Genome sequence and analysis of the tuber crop potato. Nature. 2011;475:189–95. https://doi.org/10.1038/nature10158.
4. Farber C, Mahnke M, Sanchez L, Kurouski D. Advanced spectroscopic techniques for plant disease diagnostics. A review. Trends Anal Chem. 2019;118:43–9.
5. Damodaran S, Parkin KL, editors. Fennema's food chemistry. 5th ed: CRC Press; 2017.
6. Zhu T, Jackson DS, Wehling RL, Geera B. Comparison of amylose determination methods and the development of a dual wavelength iodine binding technique. J Cereal Chem. 2008;85:51–8.
7. Mihaljev Ž, Jakšić S, Prica NB, Ćupić ŽN, Baloš MŽ. Comparison of the Kjeldahl method, Dumas method and NIR method for total nitrogen determination in meat and meat products. J Agroaliment Process Technol. 2015;21(4):365–70.
8. Osborne BG, Fearn T, Hindle PH, Hindle PT. Practical NIR spectroscopy with applications in food and beverage analysis: Longman Scientific & Technical; 1993.
9. Kim Y, Singh M, Kays SE. Near-infrared spectroscopic analysis of macronutrients and energy in homogenized meals. Food Chem. 2007;105(3):1248–55.
10. Stubbs TL, Kennedy AC, Fortuna A-M. Using NIRS to predict fiber and nutrient content of dryland cereal cultivars. J Agric Food Chem. 2010;58(1):398–403. https://doi.org/10.1021/jf9025844.
11. Berardo N, Brenna OV, Amato A, Valoti P, Pisacane V, Motto M. Carotenoids concentration among maize genotypes measured by near infrared reflectance spectroscopy (NIRS). Innov Food Sci Emerg Technol. 2004;5(3):393–8. https://doi.org/10.1016/j.ifset.2004.03.001.
12. Baranska M, Schütze W, Schulz H. Determination of lycopene and β-carotene content in tomato fruits and related products: comparison of FT-Raman, ATR-IR, and NIR spectroscopy. Anal Chem. 2006;78(24):8456–61. https://doi.org/10.1021/ac061220j.
13. Kurouski D, Van Duyne RP, Lednev IK. Exploring the structure and formation mechanism of amyloid fibrils by Raman spectroscopy: a review. Analyst. 2015;140(15):4967–80. https://doi.org/10.1039/c5an00342c.
14. Krimmer M, Farber C, Kurouski D. Rapid and noninvasive typing and assessment of nutrient content of maize kernels using a handheld Raman spectrometer. ACS Omega. 2019;4(15):16330–5. https://doi.org/10.1021/acsomega.9b01661.
15. Burlingame B, Mouille B, Charrondiere R. Nutrients, bioactive non-nutrients and anti-nutrients in potatoes. J Food Compos Anal. 2009;22:494–502.
16. Matousek P, Clark IP, Draper ERC, Morris MD, Goodship AE, Everall N, et al. Parker AW subsurface probing in diffusely scattering media using spatially offset Raman spectroscopy. Appl Spectrosc. 2005;59(4):393–400. https://doi.org/10.1366/0003702053641450.
17. Matousek P, Stone N. Development of deep subsurface Raman spectroscopy for medical diagnosis and disease monitoring. Chem Soc Rev. 2016;45(7):1794–802. https://doi.org/10.1039/C5CS00466G.
18. López-López M, García-Ruiz C. Infrared and Raman spectroscopy techniques applied to identification of explosives. TrAC Trends Anal Chem. 2014;54:36–44. https://doi.org/10.1016/j.trac.2013.10.011.
19. Bloomfield M, Andrews D, Loeffen P, Tombling C, York T, Matousek P. Non-invasive identification of incoming raw pharmaceutical materials using spatially offset Raman spectroscopy. J Pharm Biomed Anal. 2013;76:65–9. https://doi.org/10.1016/j.jpba.2012.11.046.
20. Almeida MR, Alves RS, Nascimbem LB, Stephani R, Poppi RJ, de Oliveira LF. Determination of amylose content in starch using Raman spectroscopy and multivariate calibration analysis. Anal Bioanal Chem. 2010;397(7):2693–701. https://doi.org/10.1007/s00216-010-3566-2.
21. Kizil R, Irudayaraj J, Seetharaman K. Characterization of irradiated starches by using FT-Raman and FTIR spectroscopy. J Agric Food Chem. 2002;50(14):3912–8.
22. Edwards HG, Farwell DW, Webster D. FT Raman microscopy of untreated natural plant fibres. Spectrochim Acta A Mol Biomol Spectrosc. 1997;53A(13):2383–92.
23. De Gussem K, Vandenabeele P, Verbeken A, Moens L. Raman spectroscopic study of Lactarius spores (Russulales, Fungi). Spectrochim Acta A. 2005;61(13–14):2896–908. https://doi.org/10.1016/j.saa.2004.10.038.
24. Schulz H, Baranska M, Baranski R. Potential of NIR-FT-Raman spectroscopy in natural carotenoid analysis. Biopolymers. 2005;77(4):212–21. https://doi.org/10.1002/bip.20215.
25. Edwards HGM, Farwell DW, Webster D. FT Raman microscopy of untreated natural plant fibres. Spectrochim Acta A. 1997;53(13):2383–92. https://doi.org/10.1016/S1386-1425(97)00178-9.
26. Walton A. Biopolymers, vol. 1. 1st ed. New York: Academic Press; 1973.
27. Wiercigroch E, Szafraniec E, Czamara K, Pacia MZ, Majzner K, Kochan K, et al. Raman and infrared spectroscopy of carbohydrates: a review. Spectrochim Acta A. 2017;185:317–35. https://doi.org/10.1016/j.saa.2017.05.045.
28. Zheng R, Zheng X, Dong J, Carey PR. Proteins can convert to beta-sheet in single crystals. Protein Sci. 2004;13(5):1288–94. https://doi.org/10.1110/ps.03550404.
29. Larsen KL, Barsberg S. Theoretical and Raman spectroscopic studies of phenolic lignin model monomers. J Phys Chem B. 2010;114(23):8009–21. https://doi.org/10.1021/jp1028239.
30. Cael JJ, Koenig JL, Blackwell J. Infrared and raman spectroscopy of carbohydrates. Part 4: Normal coordinate analysis of V-amylose. Biopolymers. 1975;14(1):1885–903.
31. Adar F. Carotenoids - their resonance Raman spectra and how they can be helpful in characterizing a number of biological systems. Spectroscopy. 2017;32(6):12–20.
32. Kang L, Wang K, Li X, Zou B. High pressure structural investigation of benzoic acid: raman spectroscopy and x-ray diffraction. J Phys Chem C. 2016;120(27):14758–66. https://doi.org/10.1021/acs.jpcc.6b05001.

33. Agarwal UP. Raman imaging to investigate ultrastructure and composition of plant cell walls: distribution of lignin and cellulose in black spruce wood (Picea mariana). Planta. 2016;224(5):1141–53. https://doi.org/10.1007/s00425-006-0295-z.
34. Pompeu DR, Larondelle Y, Rogez H, Abbas O, Pierna JAF, Baeten V. Characterization and discrimination of phenolic compounds using Fourier transformation Raman spectroscopy and chemometric tools. Biotechnol Agron Soc Environ. 2017;22(1):1–16.
35. Egging V, Nguyen J, Kurouski D. Detection and identification of fungal infections in intact wheat and Sorghum grain using a hand-held Raman spectrometer. Anal Chem. 2018;90:8616–21. https://doi.org/10.1021/acs.analchem.8b01863.
36. Farber C, Kurouski D. Detection and identification of plant pathogens on maize kernels with a hand-held Raman spectrometer. Anal Chem. 2018;90(5):3009–12. https://doi.org/10.1021/acs.analchem.8b00222.
37. Sanchez L, Ermolenkov A, Tang XT, Tamborindeguy C, Kurouski D. Non-invasive diagnostics of Liberibacter disease on tomatoes using a hand-held Raman spectrometer. Planta. 2020;251(3):64. https://doi.org/10.1007/s00425-020-03359-5.
38. Sanchez L, Farber C, Lei J, Zhu-Salzman K, Kurouski D. Noninvasive and nondestructive detection of cowpea bruchid within cowpea seeds with a hand-held Raman spectrometer. Anal Chem. 2019;91(3):1733–7. https://doi.org/10.1021/acs.analchem.8b05555.
39. Sanchez L, Pant S, Irey MS, Mandadi K, Kurouski D. Detection and identification of canker and blight on orange trees using a hand-held Raman spectrometer. J Raman Spectrosc. 2019;50:1875–80.
40. Sanchez L, Pant S, Xing Z, Mandadi K, Kurouski D. Rapid and noninvasive diagnostics of Huanglongbing and nutrient deficits on citrus trees with a handheld Raman spectrometer. Anal Bioanal Chem. 2019;411:3125–33.