Plant Pathol. J. 34(5) : 381-392 (2018)
https://doi.org/10.5423/PPJ.OA.02.2018.0019
pISSN 1598-2254 eISSN 2093-9280

Research Article Open Access

The Plant Pathology Journal

Detection of *Clavibacter michiganensis* subsp. *michiganensis* Assisted by Micro-Raman Spectroscopy under Laboratory Conditions

Moisés Roberto Vallejo Pérez[1]*, Hugo Ricardo Navarro Contreras[2], Jesús A. Sosa Herrera[3], José Pablo Lara Ávila[4], Hugo Magdaleno Ramírez Tobías[4], Fernando Díaz-Barriga Martínez[2], Rogelio Flores Ramírez[1], and Ángel Gabriel Rodríguez Vázquez[2]

[1]*CONACyT- Universidad Autónoma de San Luis Potosí. Álvaro Obregón #64, Col. Centro, C.P. 78000, San Luis Potosí, S.L.P. México*

[2]*Universidad Autónoma de San Luis Potosí. Coordinación para la Innovación y la Aplicación de la Ciencia y la Tecnología (CIACyT). Av. Sierra Leona #550, Col. Lomas 2a. Sección, C.P. 78210, S.L.P., México*

[3]*CONACyT- Centro de Investigación en Ciencias de Información Geoespacial A.C. Circuito Tecnopolo Norte 117, Col. Fraccionamiento Tecnopolo Pocitos, CP. 20313, Aguascalientes, Ags. México*

[4]*Universidad Autónoma de San Luis Potosí. Facultad de Agronomía y Veterinaria. Km. 14.5 Carretera San Luis Potosí, Matehuala, Ejido Palma de la Cruz, Soledad de Graciano Sánchez, C.P. 78321. S.L.P. México*

(Received on February 7, 2018; Revised on May 10, 2018; Accepted on May 31, 2018)

***Clavibacter michiganensis* subsp. *michiganesis* (*Cmm*) is a quarantine-worthy pest in México. The implementation and validation of new technologies is necessary to reduce the time for bacterial detection in laboratory conditions and Raman spectroscopy is an ambitious technology that has all of the features needed to characterize and identify bacteria. Under controlled conditions a contagion process was induced with *Cmm*, the disease epidemiology was monitored. Micro-Raman spectroscopy (532 nm λ laser) technique was evaluated its performance at assisting on *Cmm* detection through its characteristic Raman spectrum fingerprint. Our experiment was conducted with tomato plants in a completely randomized block experimental design (13 plants × 4 rows). The *Cmm* infection was confirmed by 16S rDNA and plants showed symptoms from 48 to 72 h after inoculation, the evolution of the incidence and severity on plant population varied over time and it kept an aggregated spatial pattern. The contagion process reached 79% just 24 days after the epidemic was induced. Micro-Raman spectroscopy proved its speed, efficiency and usefulness as a non-destructive method for the preliminary detection of *Cmm*. Carotenoid specific bands with wavelengths at 1146 and 1510 cm^{-1} were the distinguishable markers. Chemometric analyses showed the best performance by the implementation of PCA-LDA supervised classification algorithms applied over Raman spectrum data with 100% of performance in metrics of classifiers (sensitivity, specificity, accuracy, negative and positive predictive value) that allowed us to differentiate *Cmm* from other endophytic bacteria (*Bacillus* and *Pantoea*). The unsupervised KMeans algorithm showed good performance (100, 96, 98, 91 y 100%, respectively).**

Keywords **:** chemometrics, epidemiology, KMeans, LDA, PCA

Handling Associate Editor **:** Oh, Chang-Sik

*Corresponding author.
Phone) +52 (444) 826 2300, FAX) +52 (444) 826 8410
E-mail) vallejo.pmr@gmail.com

The tomato crop (*Solanum lycopersicun* L.) is a domesticated species belonging to the *Solanaceae* family. Taxonomically, it is located in the *Lycopersicon* section alongside 13 related wild species, which can be found on the western coast of South America (Ecuador, Peru, and Chile), this region is considered to be their center of origin; nevertheless, domestication of cultivated tomato occurred

in México (Knapp and Peralta, 2016). The global importance of the tomato crop resides in the fact that its fruits are an important source of vitamins and minerals in the human diet, as they are frequently consumed both fresh and processed, amounting to 120 million tons produced annually worldwide (Colvine and Branthôme, 2016; FAO, 2017). Corporate farming is supported by cultivating high-performing varieties, which also enhance desirable traits to consumers and tend to be grown in monocropping systems on broad swaths of farmlands. Nevertheless, even with the availability of phytopathogen resistant cultivars, the emergence of new pathogen strains and antibiotic-resistant variants constitute risk factors that may potentially prompt an epidemic outbreak that could undermine food security (Fletcher et al., 2006).

Evaluating and predicting pathogen effects on cultivars represent a continuous challenge due to the low genetic diversity of crops, which increases the potential for pathogen spreading rate and crop damage. Thus, it is necessary to advance in the development of new technologies and infrastructure to execute more-robust surveillance mechanisms, appropriate sampling protocols, and disease diagnosis techniques (Fletcher et al., 2006). Currently, there are many types of methods for rapid and accurate identification of dangerous biological agents, these can be further divided into conventional, and modern technologies, and there are also integrated and automated diagnostic systems for detection and identification of microorganisms (Mirski et al., 2014). Raman spectroscopy belongs to the family of vibrational spectroscopic techniques, the instrument creates specific Raman spectral fingerprints by recording the molecular vibrations of the cellular compounds, and it is currently advertised as a hot and ambitious technology that has all of the features needed to characterize and identify bacteria, because Raman fingerprint represents the complete information of the biochemistry of the cell (Ashton et al., 2011; Lorenz et al., 2017). A modern Raman microscope allows to analyze bacteria at single-cell level and the application of this approach can be used to readily and specifically detect plant pathogens (Gan et al., 2017). Moreover, this technique can even be employed for functional analyses of microbial communities, but its biggest challenge is that spontaneous Raman signals are naturally weak, especially at single cell levels (Wang et al., 2016).

The present research is focused on the study of *Clavibacter michiganensis* subsp. *michiganensis* (*Cmm*), which causes the bacterial canker of tomato (BCT), a quarantine-worthy pest included on the list of regulated plant pests in Mexico and the United States of America, among other countries (EPPO, 2016). *Cmm* bacterium spreads by seeds and it is highly contagious. At the moment, there are no genetic materials showing resistance or tolerance to the *Cmm* bacterium (Sen et al., 2013). BCT occurrence is usually erratic, the disease is asymptomatic at its early stages, and the compatible interaction between *Cmm* and tomato is an extremely complex and multi-faceted biological system, one of whose facets is the level of genetic variability in each organism (Martínez-Castro et al., 2018).

Therefore, the objective of this research involves evaluation under controlled conditions the utility of Raman spectroscopy (532 nm λ laser) technique to reduce the time for bacterial recognition by using a preliminary detection of *Cmm* colonies through its Raman spectrum fingerprint and chemometric analyses in stationary laboratory conditions.

Materials and Methods

Experimental establishment and plant inoculation. The experiment was conducted with Ramses F1 (Harris Moran®) tomato plants (*S. lycopersicun* L.). The trial was conducted under net-house conditions in the CIACyT-UASLP located at coordinates 22.150715 N, -101.025097 W, and elevation of 1850 MASL. We used a completely randomized block experimental design composed of four blocks (rows) and 13 plants per block (repetitions). All the plants were kept isolated using anti-aphid mesh (net-house) in a fertirrigation system in plastic growing bags with a peat-based substrate (Premier® Horticulture Inc. Canada), and agronomic practices were implemented (irrigation, fertilization, and crop management) in accordance to recommendations for the zone (Jasso et al., 2012).

After 5 weeks of growth cycle, nine randomly selected plants were inoculated with *Cmm* bacteria. The *Cmm* pathogenic isolate was supplied by the National Center of Genetic Resources (CNRG) under the auspices of the National Forestry, Agricultural, and Livestock Research Institute (INIFAP). The selected method to inoculate the plants drew on Sen et al. (2013) with some modifications. The inoculum was prepared with *Cmm* isolate (supplied by CNRG-INIFAP) grown on Luria-Bertani (LB) liquid medium at 28°C at 150 rpm until reached an optical density of 0.2 at 600 nm (Multiskan GO™, Thermo Fisher Scientific Inc.) corresponding to about 10^8 cfu/ml. A piece of sterilized cotton soaked with 100 µl of the bacterial solution was held (Parafilm®) upon 2 superficial incisions of 2 mm long located at stem section between the second and third true leaf. For the negative controls, three plants were inoculated with sterile distilled water. Finally, the plants were wrapped in polyethylene bags to raise the relative humidity over a 24 h period. After this time period, the *Cmm* inoculated

plants were placed within the original plant population that conformed the experiment to induce contagion to neighboring plants and the disease epidemiology was evaluated.

Evaluating the severity and incidence of BCT. Disease incidence and severity were evaluated daily for 40 days until the plants reached 14 weeks of age. Severity was evaluated using the scale proposed by Sen et al. (2013), consisting of five categories: no symptoms = 0; one leaf with wilting symptoms = 1; more than one wilted leaves, less than 50% of the leaves with wilting symptoms = 2; between 50% and 75% of leaves wilted = 3; more than 75% of leaves wilted but less than 100% = 4; the entire plant wilted and dead = 5. Relative incidence (Inc-R) was defined as the number of plants with symptoms of the disease observed only on the date of evaluation, while cumulative incidence (Inc-C) amounted to the total number of plants that showed symptoms before and during the evaluation. The spatio-temporal pattern of the epidemic was analyzed using zone mappings under the interpolative geostatistical analysis method in the Surfer® program vers. 14, represented by three-dimensional and contour maps. Additionally, Morosita aggregation indices were calculated pursuant to Campbell and Madden (1990), and the area under the disease progress curve (AUDPC) was calculated using the Guillén-Sánchez et al. (2003) trapezoid method. Temperature and precipitation weather variables were monitored using a Davis Vantage Pro2™ (Davis Instruments, Hayward USA) weather station.

***Cmm* isolation and molecular identification.** The phytosanitary analysis was performed on all plants than conformed the experiment at 14 weeks of age to determine the presence of *Cmm* in plant's vascular tissue. The procedure consisted of cutting 10 cm of stem located between the second and third leaf, which were then washed with soap and running water, superficially disinfected with 70% alcohol (2 mins) and 1.5% sodium hypochlorite (1 min), and rinsed with sterile deionized water. Later, the stems were dissected and the internal vascular stem tissue with brown coloring was extracted and submerged in 5 ml of sterile distilled water for 15 mins. The suspended mixing allowed diffusion of bacteria out of plant tissue and the bacteria were grown in Nutritive Agar medium (Bioxon™) and incubated at 28°C for 72 h. Bacterial colonies isolated were purified and classified pursuant to their colony morphology and Gram staining (EPPO, 2016).

The bacterial genomic DNA was extracted using lysis buffer (Tris Base 50 mM, EDTA 50 mM, and SDS 3%), in accordance with the procedure described by Sambrook and Russell (2001) beginning with bacterial isolates that had been previously augmented in B-King liquid media for 24 h at 28°C (EPPO, 2016). The PCR amplification process was done using the genomic DNA of the genetic region 16S of ribosomal DNA with oligonucleotides F27 (5'-AGAGTTTGATCMTGGCTCAG-3') and R1492 (5'-TACGGYTACCTTGTTACGACTT-3') under the following conditions: 95°C for 5 mins, followed by 35 cycles at 95°C for 45 s, 55°C for 45 s and 72°C for 90 s, and a final extension at 72°C for 5 mins (Monciardini et al., 2002). PCR products of expected sizes were purified and sequenced. The taxonomic assignment of partial sequences larger than 1 Kb of 16S rDNA genes was performed with the classification service implemented in the SINA Alignment Service using the SILVA database with default parameter settings, the minimum identity with the query sequence was adjusted to 0.90 (Pruesse et al., 2012; Quast et al., 2013). The BLASTn algorithm was used to search the NCBI GenBank database (Benson et al., 2010) to confirm taxonomical assignment of sequences.

Detection of *Cmm* assisted by Micro-Raman Spectroscopy. The preliminary detection of *Cmm* through its Raman spectrum fingerprint was done under stationary laboratory conditions. The analysis was carried out according to Paret et al. (2010, 2012) with some modifications. All the bacterial pure isolates (colonies) obtained were grown individually in LB liquid media (5 ml) and incubated at 28°C and 200 rpm until reached an optical density of 0.1 at 600 nm. The bacteria was individually harvested by centrifugation (14,000 rpm for 3 mins at 8ºC) and washed once in 0.85% NaCl and twice using sterile deionized water. Cells washed were resuspended in 100 µl of sterile deionized water and the optical density was adjusted to 0.1 at 600 nm (Paret et al., 2010). For the Micro-Raman analysis, a polished mirror aluminum sheet of 20 × 30 mm^2 and 0.2 mm in thickness, with made micro-cavities of 300 µm diameter was used as a substrate to obtain the Raman spectra (Misra et al., 2009). The aluminum substrate was cleaned with methanol and dried, 3 µl of each processed bacterial suspensions were placed in a separated micro-cavity (4 repetitions per pure isolate) and left to dry for 30 minutes. Negative control samples were obtained without the addition of bacteria (sterile deionized water). The corresponding Raman fingerprint was obtained from the bacterial sample deposited on the micro-cavity using a Raman Confocal Horiba XploRA ONTM™ (Horiba Scientific, Ltd.) microscope equipped with a 532-nm (HeNe) under the following conditions: spectral range of 100-2000 cm^{-1}, acquisition time 10 s to avoid sample damaging, laser power ≈ 20 mW, 1200 gr/

mm grating, slit 100 µm, hole 300 µm, 10X lens (microspot with 10 µm diameter) and CCD cooled detector with Peltier system. The CDD resolution is 2 cm^{-1}.

Spectral Preprocessing and Computational Analysis. Raman spectra datasets obtained from different bacterial colonies were identified and stored in comma separated values (CSV) format, then, they were preprocessed to eliminate background fluorescence by subtracting a fifth-order polynomial from the original spectra and normalizing them to the area under the curve measure for the entire spectral range using the Vancouver Raman Algorithm program (Zhao et al., 2007). To perform the computational operations on our data, we wrote the corresponding preprocessing and calling routines on Python programming language. We employed the Scikit-Learn (Pedregosa et al., 2011) library, which is a collection of machine learning libraries used in several fields of research. We used Principal Components Analysis (PCA) (Wold et al., 1987) for feature selection, then, we applied Linear Discriminant Analysis (LDA) and KMeans clustering algorithm as classifiers.

PCA is a widespread method employed for dimensionality reduction, multivariate correlation analysis and model quality evaluation. It is based on the projection of a set of n data samples $\{X_i\}$ with $X_i = (x_i, ..., x_{im})$ from a m-dimensional space, onto a k-dimensional feature space generated by a basis of orthogonal eigenvectors $\{\hat{e}_1, ..., \hat{e}_k\}$ corresponding to the k largest eigenvalues $\{\lambda_1, ..., \lambda_k\}$ of X_i (Abdi and Williams, 2010). PCA can be used to increase class separation among elements in a dataset by reducing correlation of the projected representation as each λ_i is a measure of the variance along the respective direction $\hat{e}_i$ (Jolliffe, 2002). PCA projection matrix W is obtained mainly through Singular Value Decomposition (SVD) on modern software (Abdi and Williams, 2010). In this study we employed PCA to obtain a previous feature selection for the classifiers used in order to improve their performance. The number of components N_{comp} used for PCA was chosen in such way that it gave optimal results for all methods, so their performance could be compared in a consistent manner. To this end, a sequential execution of the algorithms used in this work was performed and the metrics were recorded for each execution, iterating over the range of $N_{comp} \in [1, 50]$. We restricted the number of testing components to this range because eigenvalues λ_i of our dataset became very close to zero for $i > 50$, and they could not be accurately represented by computer double precision floating point data types, making PCA numerically unstable beyond such range. Once N_{comp} was determined, we used this value for all classification algorithm executions. Metrics taken from algorithms using a previous step of feature selection with PCA, were all using the same number of components N_{comp}.

LDA is a supervised feature selection and classification method, which simultaneously maximizes the distance between the mean classes while minimizing the variance within each class. Considering a set of samples $\{X_i\}$ of m-dimensional vectors partitioned into $N = 2$ subsets $\{X_i\} = A \cup B$, $A \cap B = \varnothing$, $A, B \neq \varnothing$ with maximally separate means μ_A and μ_B and minimal within class variance (Xanthopoulos et al., 2013). For the case $N > 2$ classes, LDA is generalized to Multiclass LDA by the use of multivariate analysis of variance. Although LDA can be used for both feature selection and classification, due to the high dimensionality of the Raman spectrum samples, we use it in here only as a classifier. We divided the dataset into two subsets for training and testing executions, each one containing respectively 70% and 30% of the original data in a random way.

Additionally, the KMeans clustering is an efficient unsupervised clustering algorithm often used for the large number of samples and features (Izenman, 2008). In the KMeans algorithm, an initial set $X = \{X_i\}$ of n samples, is partitioned into $C = \{C_1, ..., C_k\}$ classes. KMeans tries to minimize the distance of an element X_i to the centroid μ_j of the cluster C_j at which X_i belongs, by iteratively reallocating the μ_j positions, the resumed steps for KMeans are describe in Hartigan and Wong (1979) and Jolliffe (2002). To perform the unsupervised classification on the Raman spectra dataset, we executed 20 repetitions of the KMeans algorithm with different random initializations.

Finally, to evaluate the performance of the classifiers employed in this paper (PCA, LDA, PCA+LDA and PCA+KMeans), we compute several metrics used for bi-

Table 1. Quality metrics used

Validation and quality tools	Equations
Sensitivity	$\frac{TP}{TP + FN} \times 100$
Specificity	$\frac{TN}{TN + FP} \times 100$
Positive predictive value (PPV)	$\frac{TP}{TP + FP} \times 100$
Negative predictive value (NPV)	$\frac{TN}{TN + FN} \times 100$
Accuracy	$\frac{(\text{Sensitivity} + \text{Specificity})}{2}$

TP = True Positive, FP = False Positive, TN = True Negative, FN = False Negative.

nary classification assessment (Siqueira and Lima, 2016; Velez et al., 2007), that includes metrics corresponding to Sensitivity (SENS), defined as the confidence that a positive result for a sample of the labeled class is obtained; the Specificity (SPEC), which is the confidence that a negative result for a sample of non-labeled classes is obtained; the Positive Predictive Value (PPV) measuring the proportion of positives that are correctly assigned, the Negative Predictive Value (NPV) which measures the proportion of negatives that are correctly assigned, and finally the Accuracy (ACC), that is the proportion of all tests that are correctly classified (Santos et al., 2017). Table 1 summarizes these equations.

To have an insight into the performance behavior of each algorithm we computed the correlation $r_{i,k}$ of the original spectral band Y_i at the Raman spectrum of the samples to the k-th principal component by the equation:

$$r_{i,k} = \frac{\lambda_k \alpha_{ik}}{var(Y_i)^{1/2}}$$

Where α_{ik} is the j-th coefficient of the eigenvector $\hat{e}_k$ (Jolliffe, 2002).

Results

Incidence and severity of the bacterial canker of tomato. The plants analyzed were classified into one of three categories due to the disease behavior: (i) plants inoculated with *Cmm* (In+), (ii) plants positively infected by contagion with *Cmm* (Co+), and (iii) negative asymptomatic plants (N–). After forty-eight to seventy-two h of *Cmm* inoculation, the infected tomato plants (In+) began to show disease symptoms of bacterial canker of tomato (BCT), the entire leaves became wilted and the stem tissue showed discoloration at the inoculation point; however, some plants later became asymptomatic and the inoculation site on the stem evolved into a corky canker.

The relative incidence (Inc-R) behaved dynamically in the population, with a waxing and waning number of symptomatic plants. The experiment began with 9 plants artificially infected with *Cmm* (In+), but 23 days later, 40% of the population (21 plants) simultaneously showed some degree of disease severity (In+, Co+); nevertheless, the cumulative incidence (Inc-C) observed over the 40 days of evaluation amounted to 79% (38 plants), peaking at 24 days after the epidemic was induced. BCT disease progress curve of the Inc-C behaved logarithmically ($R^2 = 0.859$) (Fig. 1).

The severity of the symptoms associated with *Cmm* also behaved dynamically, at both the individual and population levels, as they were displaying rising and falling values over the growth cycle. The interpolation maps confirmed the observations made throughout the experiment (Fig. 2). The scale value 5 (dead plant) was observed solely in two artificially-inoculated plants (In+), but the 7 remaining inoculated plants (In+) exhibited values ranging from

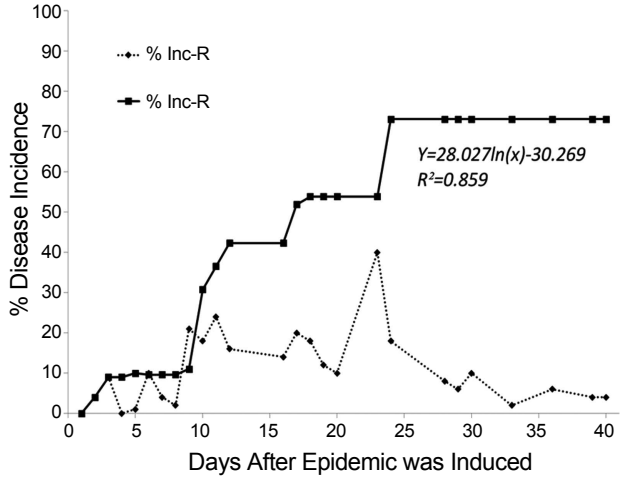

Fig. 1. Temporal progress curves of relative incidence (Inc-R) and cumulative incidence (Inc-C) of tomato plants infected with *C. michiganensis* subsp. *michiganensis*.

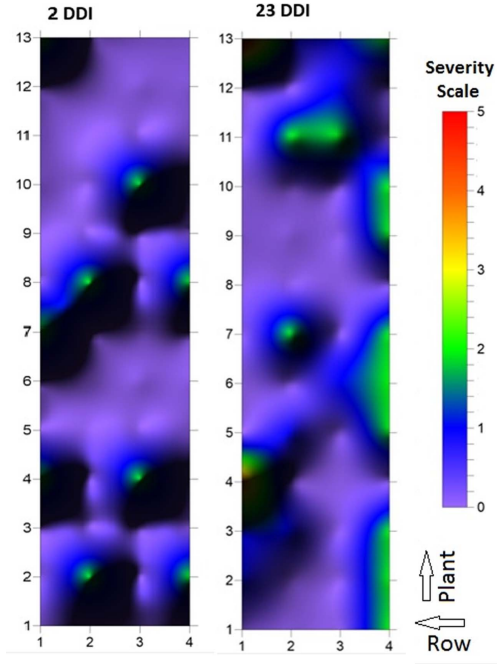

Fig. 2. Interpolation maps of disease severity: 2 days (Morisita Index: < 1) and 23 days (Morisita Index: 1.7) after the epidemic was induced (DDI) with *C. michiganensis* subsp. *michiganensis* in tomato plants.

0, 1, 2, and 3 on the scale. Comparing the area under the disease progress curve (AUDPC), the artificially-infected plants (In+) exhibited statistically higher values ($P \leq 0.05$) (AUDPC = 25) than the contagion-infected plants (Co+) (AUDPC = 6.3), and the last ones (Co+) were generally those neighboring the artificially inoculated plants (In+). Only 11 plants (21%) were symptom-free (N–) 40 days after the epidemic was induced, and their AUDPC values were practically null. The spatio-temporal pattern, two days after the epidemic was induced (DDI) resembled the structure of a uniform distribution (Morisita Index: < 1), but later evaluations showed an aggregated pattern (Morisita Index: 1.7) (Fig. 2) (Campbell and Madden, 1990). It is worth to note that the highest incidence occurred in rows 3 and 4, where runoff converged during irrigation, and that incidence was most evident 23 days after the epidemic was induced (Fig. 2).

Molecular Identification of *Cmm*. The bacterial colony morphology of the *Cmm* isolate provided by CNRF-INIFAP and used in the experiment could be characterized as follows: Gram (+), yellow, circular, smooth-edged, flat and elevated surface, and creamy texture. The phytosanitary analysis performed at 14 weeks of age in plant population to determine the presence of *Cmm* in the vascular tissue, allows the isolation of different bacterial colonies from the symptomatic and asymptomatic tomato plants (In+, Co+) and displaying different colony morphologies. Strains with colony morphologies similar to the *Cmm* isolate (supplied by CNRG-INIFAP) were obtained from both inoculated (In+) and contagion infected (Co+) plants and they showed the same colony morphology of previously described *Cmm* CNRF-INIFAP isolate. The molecular 16S rDNA analysis of reisolated bacterial strains were classified as *Clavibacter michiganensis* subsp. *michiganensis* (NCBI Accession Numbers: HQ144239.1, KR922121.1, HQ144230.1, HQ144239.1, KR922121.1, HQ144230.1). Additionally, other endophytic bacteria were isolated and molecularly identified, turning out to be the genera of *Bacillus* sp. (NCBI Accession Number: HM566983.1) and *Pantoea* sp. (NCBI Accession Number: MF352035.1, KU933320.1). Therefore, in our study we confirmed the infection of tomato plants with the virulent C*mm* CNRF-INIFAP isolate.

Spectral Raman features. The Micro-Raman instrumentation allows to obtain the spectral signature of *Cmm* and the acquisition time (10 s) does not cause any bacterial cell damage (Fig. 3A). The characteristic Raman spectrum of the *Cmm* CNRF-INIFAP isolate was characterized by wavelength peaks at 944, 992, 1146, 1179, 1254, 1435,

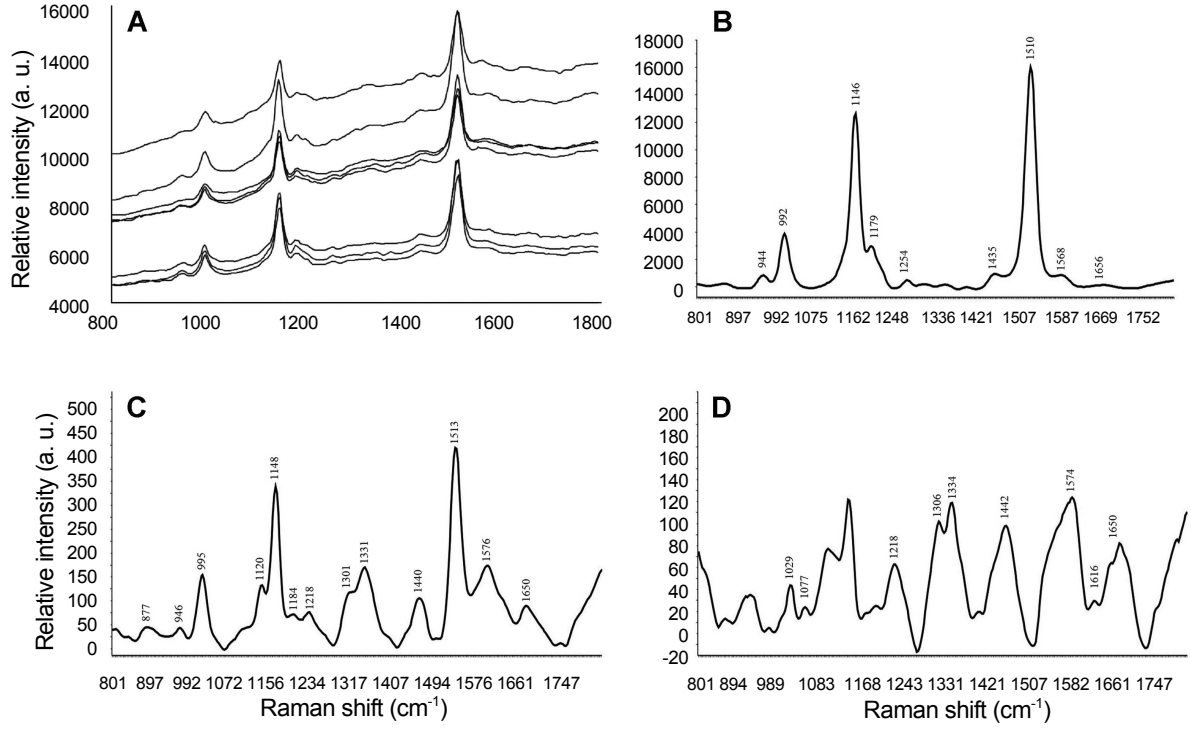

Fig. 3. Raman spectra obtained at 532 nm λ corresponding to the bacteria: (A) *Clavibacter michiganensis* subsp. *michiganensis*, raw spectra (B) *C. michiganensis* subsp. *michiganensis* preprocessed spectra (average 4 repetitions), (C) *Pantoea* sp., and (D) *Bacillus* sp.

Table 2. Peak positions of the Raman bands of the bacteria isolated in this work. Spectra were recorded in the 800-1800 cm^{-1} region, with a 532 nm λ laser

Bacteria	Wavenumbers (cm^{-1})	Spectrum
Cmm	944, 992, 1146, 1179, 1254, 1435, 1510, 1568, 1656	b
Pantoea sp.	877, 946, 995, 1120, 1148, 1184, 1218, 1301, 1331, 1440, 1513, 1576, 1650	c
Bacillus sp.	1029, 1077, 1218, 1306, 1334, 1442, 1574, 1616, 1650	d

1510, 1568, and 1656 cm^{-1} (Fig. 3B). The respective peaks varied in relative intensity, but among them, bands at 1146 and 1510 cm^{-1} stand out, which are the vibrational modes of the C=C and C−C bonds (stretching) of the trans isomer in the carotenoids (Jehlička and Oren, 2013). The 992 cm^{-1} band, matching vibrational mode A_{1g} of the benzene ring (Chen et al., 2007), and band 1179 cm^{-1}, associated with the saccharides (Athamneh and Senger, 2012), generally found as components of the cell wall in *Cmm*, were observed at smaller intensity. Bands with relatively weak intensity were found at wavelengths 944 cm^{-1}, associated with phenyl ring CH and COH (bend) (Maiti et al., 2013), band 1254 cm^{-1}, associated with mode ν_{asym} (O-P-O) (Movileanu et al., 1999), band 1435 cm^{-1} associated with vibration C-C stretch (D-Band) and the band 1568 cm^{-1} characteristic of carbon G band (C-C) (Malard et al., 2009) and, finally, the band 1656 cm^{-1}, associated to ν(C=O) of the amide I (Gelder et al., 2007). Raman spectra of *Pantoea* sp. (Fig. 3C) and *Bacillus* sp. (Fig. 3D) were different from those observed in *Cmm* (Table 2), and their Raman spectra match those previously reported by other authors (Polisetti et al., 2016). It is important to keep in mind than during routine analysis, control samples (positive and negative) should always be included, because slight instrumental displacements (± 5 cm^{-1}) may happen during spectral measurements.

Computational Analysis. Supervised and unsupervised computer classification analyses were conducted on 75 bacterial isolates, of which 22 were *Cmm* and 53 were endophyte bacteria (*Bacillus* sp. and *Pantoea* sp.). To automatically classify our spectra dataset we proceeded in the following way: we applied the unsupervised KMeans classifier directly to the Raman spectra data (KMeans), we also classified the spectra by previously performing dimensionality reduction via PCA, and selected the first 12 principal components of each spectrum (PCA+KMeans). For the supervised LDA method we also applied the classifier directly over spectral vector (LDA) and we also use it in conjunction with PCA (PCA+LDA).

The number of principal components for feature selection than produced simultaneously an optimal performance

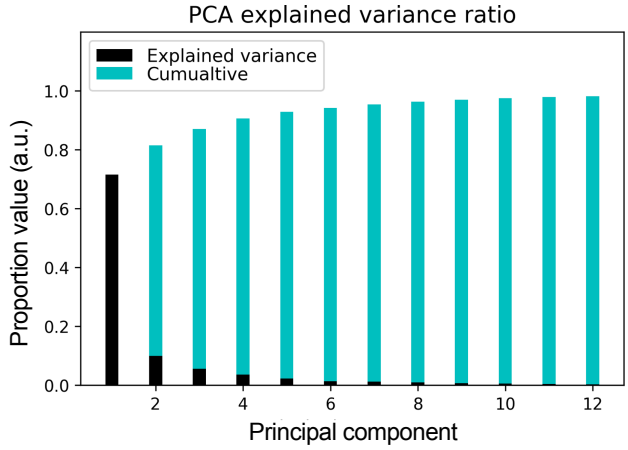

Fig. 4. The number of principal components for feature selection and classification of Raman spectra with supervised and unsupervised algorithms.

for both PCA+KMeans and PCA+LDA algorithms was searched as described in the methodology section. The value found was N_{comp} = 12. Naturally, the optimal values found on each case were very different, even though they were using the same value of N_{comp} at the PCA feature preselection stage. The explained variance for each of the first 12 principal components of our dataset, along with their corresponding cumulative values are depicted in Fig. 4, which shows that these components account for 98.2% of the total variance.

We see that the KMeans classifier can differentiate Raman spectral signatures of the bacteria analyzed (*Cmm vs* Endophytic) when it is applied directly to them with SENS of 100%, SPEC of 96%, ACC of 98%, PPV of 91% and

Table 3. Performance metrics of classifiers (percentage)

(%)	KMeans	PCA+ KMeans	LDA	PCA+LDA
SENS	100	95	-	100
SPEC	96	13	-	100
PPV	91	31	-	100
NPV	100	87	-	100
ACC	98	54	-	100

SENS = Sensitivity, SPEC = Specificity, PPV = Positive Predictive Value, NPV = Negative Predictive Value, ACC = Accuracy.

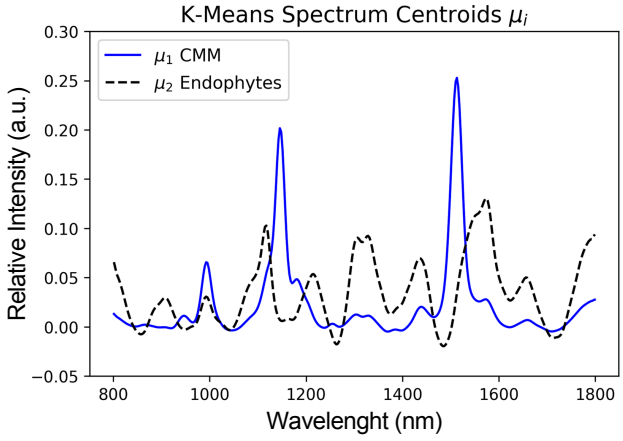

Fig. 5. Class centroids obtained by KMeans classifier for Raman spectrum samples of *C. michiganensis* subsp. *michiganensis* and Endophytic bacteria.

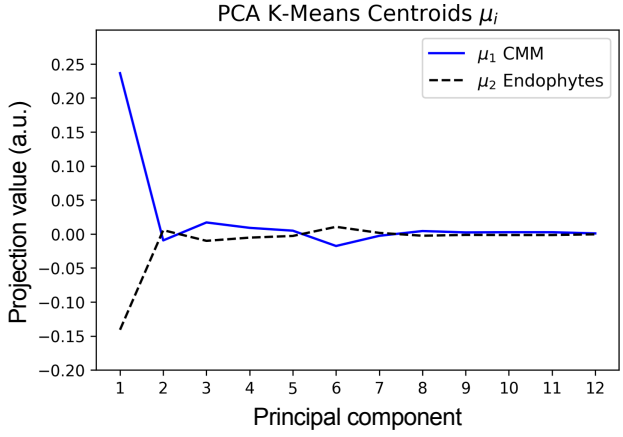

Fig. 6. Class centroids obtained by KMeans classifier for principal components of Raman spectrum samples of *C. michiganensis* subsp. *michiganensis* and Endophyte bacteria.

NPV of 100%. On the other hand, Table 3 shows an important loss on performance metrics for KMeans when used along with PCA dimensionality reduction (PCA+KMeans).

The LDA classifier did not manage to distinguish the two groups (*Cmm vs* Endophytic) when applied directly to the spectral samples. Therefore, to reduce the ratio of the number of dimensions to number of samples, we opted to use feature selection via PCA. The execution of the PCA+LDA method returned an exact classification and the performance evaluation were of 100% for SENS, SPEC, ACC, PPV and NPV, without producing any false positive (FP) or false negative (FN). On our tests, PCA+LDA can achieve better levels of specificity and accuracy than KMeans. We also observed that PCA+LDA has a low probability of overfitting when applied to identify Raman spectra of *Cmm* against endophyte bacteria signatures, as in

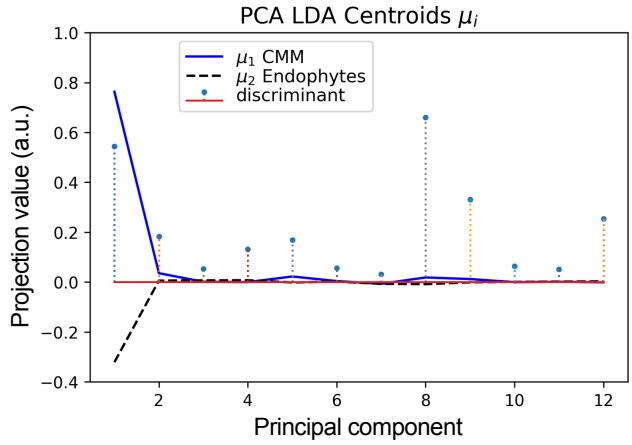

Fig. 7. Class centroids obtained by LDA classifier for principal components of Raman spectrum samples of *C. michiganensis* subsp. *michiganensis* and Endophyte bacteria. Stems represent absolute values of components of the eigenvector used by LDA for class separation.

Table 4. PCA correlations to spectral bands associated to key compounds (bold-typed) to differentiate *C. michiganensis* subsp. *michiganensis* bacteria (fingerprint)

Wavenumbers (cm^{-1})	Band Correlations		
	PC1	PC8	PC9
840	0.08898	0.27657	0.12938
912	0.79354	0.07651	0.09566
944	**0.74578**	**0.16238**	**0.05213**
992	**0.83523**	**0.03110**	**0.24548**
1070	0.59842	0.23792	0.07671
1146	**0.99055**	**0.04614**	**0.03336**
1254	**0.55826**	**0.20966**	**0.08965**
1332	0.97574	0.01291	0.06342
1460	0.45860	0.31558	0.19149
1510	**0.99045**	**0.09225**	**0.07836**
1568	**0.95005**	**0.08255**	**0.01474**
1656	**0.88012**	**0.15466**	**0.24434**
1745	0.53756	0.25293	0.03216

Values shown are for the three principal components with the largest LDA discriminant weights.

our experiments the testing and training datasets returned exactly the same classification.

The class centroids (means) found by KMeans applied directly to Raman spectra are displayed in Fig. 5 and the Fig. 6 illustrates the corresponding centroids when KMeans is applied to the principal components extracted by PCA. Note that the values shown do not belong to the principal components but to the projection of the class centroids onto such components. Fig. 7 shows the class centroids found

by LDA applied over the same principal components and stand out the PCA1, PCA8 and PCA9; additionally, the absolute values of the components of the eigenvector estimated by LDA are shown in Table 4 in order to have visual comparison of the components weights used by the LDA classifier.

Discussion

The pathogenic relationship between the tomato (*S. lycopersicum*) and the *Cmm* bacterium arose before the crop was domesticated. In this evolutionary process, the bacterium adapted itself to the vascular tissue as an endophyte, subsequently it evolved as a pathogen (Bentley et al., 2008). As such, the bacterium can systematically infect the plant through the vascular tissue, and in the early stages, the pathogen behaves like a biotrophic organism, inducing latent asymptomatic infections (Eichenlaub and Gartemann, 2011; Sharabani et al., 2013).

In this research, the *Cmm*-inoculated plants began to show disease symptoms from 48 to 72 h after inoculation; however, the evolution of the incidence and severity varied over time in the plant population. The plants inoculated at the beginning of the experiment displayed more severe symptoms pursuant to the area under the disease progress curve (AUDPC), and several even died (22%). Contrary to that, the plants infected by contagion remained asymptomatic, in accordance with previous reports, mentioning that plant age can lead to asymptomatic infections, until the plant reaches the productive stage and the seeds of the fruits become *Cmm*'s principal means of spreading (Eichenlaub and Gartemann, 2011). The concentration of the bacterium at the time of the infection is also determinant, because in concentrations below 10^4 colony-forming units (CFU), the movement of the bacterium is limited, but at 10^8 cfu, the bacterium moves up and down in the vascular tissue of the plant from the point of inoculation (Xu et al., 2010).

The contagion between plants advanced rapidly, reaching 79% just 24 days after the epidemic was induced. Chang et al. (1991) reported similar results, remarking that the speed of the epidemic is a function of the initial disease incidence. The irrigation water is likely to be the principal mean by which *Cmm* spreads (Xu et al., 2010), because during the experiment the highest disease incidence was found in rows 3 and 4. Additionally, the plants infected by contagion, generally were located in proximity to the inoculated plants, due to close contact between them (Xu et al., 2010), leading to clustered (foci) distribution patterns 23 days after the epidemic was induced. This outcome agrees with Kawaguchi et al. (2010) results, who mentioned that in addition to irrigation water, infected material (seed and/or seedling), leaf-pruning, and disbudding practices all condition the directionality of infections and how they cluster. Therefore, the opportune detection of *Cmm* symptomatic plants, allows defining the hotspots for bacterial dispersion with a preponderance of aggregated spatial pattern. In addition, the cultural conditions (irrigation, pruning directionality and environmental conditions) will conditioned the samplings protocols design to determine the cumulative incidence of the BCT disease. It is worth mentioning that on cool (26-28°C), cloudy, and rainy days, the unilateral wilting symptoms and marginal leaf necrosis were exacerbated, as these conditions are likely to boost bacterial growth inside the plant (Sen et al., 2015).

The complex parasitic relationship between *S. lycopersicum* and *Cmm* renders many of the epidemiological monitoring mechanisms moot (Sen et al., 2013), and because *Cmm* is considered a quarantine-worthy pest in various countries and economic regions, early detection becomes indispensable in order to implement timely measures to eradicate the pest and thereby to prevent it from spreading. The Micro-Raman spectroscopy used to differentiate *Cmm* bacteria isolated from symptomatic and asymptomatic plants stood out most for its speed, because the spectral analysis of the bacterial cells is done in a matter of minutes. The phytosanitary analysis, the most time spend is used to isolate the bacteria from the plant tissue; moreover, the spectral traits of *Cmm* allows to differentiate them from other endophytic bacteria, even those with similar colony morphology (Paret et al., 2010). Bands 1146 and 1510 cm^{-1} are the primary markers that set *Cmm* to separate from the rest of the bacteria that are found in the vascular tissue and these bands are associated with vibration modes of the carotenoids that are a structural part of the *Cmm* membrane (Saperstein et al., 1954).

The chemometric analysis and performance evaluation demonstrated that PCA+LDA was the best classifier algorithm with 100% of SENC, SPEC, ACC, PPV and NPV; it was followed by KMeans algorithm applied directly on spectral band values. The classifiers PCA+KMeans algorithm and single LDA alone used as a classifier did not show good performance. To explain why PCA+LDA performed much better than PCA+KMeans, we want to point out that LDA employs discriminant coefficients that minimize intra-class variance while maximizing inter-class separation. For our samples, we can see in Fig. 7 that LDA assigns the largest discriminant coefficients to the principal components PC1, PC8, and PC9. The local peaks of principal component correlations to original spectral band and their numerical values are shown in Table

4. Some correlation peaks correspond to wavenumbers of vibrational modes of key compounds linked to presence of *Cmm*. Specifically the bands at 944, 992, 1146, 1254, 1510, 1568, and 1656 cm^{-1}, which are marked in bold types, but the higher correlation values corresponded to the carotenoid bands (1146 and 1510 cm^{-1}) that distinguish *Cmm* bacteria and those are considered as unique markers (Paret et al., 2010). We can also see that there are several other wavenumbers of very week intensity having correlation peaks for principal components PC1, PC8, and PC9; however, due to Raman signals are naturally weak, the use of nanoparticles to enhance signals (Surface Enhance Raman Spectroscopy) (Wang et al., 2016) will probably be required in future work.

Such spectral features helped PCA+LDA to achieve a performance of 100% on the classification metrics employed. On the other hand, PCA+KMeans do not assign discriminant weights to principal components, which implies that it gives all the principal components exactly the same importance, even if they are not correlated with specific bands associated to *Cmm*, thus metrics show a poorer performance for PCA+KMeans with respect to PCA+LDA (Table 3). Note that for PCA+LDA, being a supervised method, it is required to separate the dataset into training and testing samples, and to have a previous knowledge of the corresponding labels for the training set. While PCA+KMeans losses accuracy at the identification of *Cmm* based on Raman spectrum of samples, it still important to consider it, because it has the advantage of being an unsupervised method. This characteristic makes PCA+KMeans convenient for circumstances when there are very few samples for the separation into training and testing sets, or when there are no previous knowledge about the samples, which makes impossible to assign labels for a training set in such cases.

The development of readily available, reliable and fast procedures to detect phytopathogens could raise the efficiency of sampling and search procedures (Fletcher et al., 2006; Vallejo et al., 2016). Diagnosis by way of conventional procedures can take one day, but it generally takes a lot longer, depending on the bacterium in question; furthermore, in the event of an epidemic when it is necessary to process a lot of samples at once, conventional procedures can be unaffordable. Thereby the importance of the present study as we demonstrated the usefulness of the Micro-Raman spectroscopy as a fast and efficient method for the preliminary identification of *Cmm*, whose detection may otherwise be difficult in culture media due to the abundant growth of saprophytes (endophytic bacteria), in accordance with Paret et al. (2010).

Acknowledgments

The present research was supported by The National Council of Science and Technology (CONACYT) through the Cátedras-CONACyT program. The authors would like to thank to the company Agriestrella S.A de C.V. and his research manager Ing. Tomas Alejandre Aumada for facilitating the strain of *Clavivacter michiganensis* subsp. *michiganensis* through the National Center of Genetic Resources (CNRG-INIFAP). This work was partially supported by CONACYT-SEP project number 236066. The technicians Lic. Álvarez Preciado L.G., and Q.F.B. Estrada Loredo Sarahi J., thanks for their support during the laboratory activities.

References

Abdi, H. and Williams, L. J. 2010. Principal component analysis. *WIREs Comp. Stat.* 2:433-459.

Ashton, L., Lau, K., Winder, C. L. and Goodacre, R. 2011. Raman spectroscopy: lighting up the future of microbial identification. *Future Microbiol.* 6:991-997.

Athamneh, A. I. M. and Senger, R. S. 2012. Peptide-guided surface-enhanced Raman scattering probes for localized cell composition analysis. *Appl. Environ. Microbiol.* 78:7805-7808.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. and Sayers, E. W. 2010. GenBank. *Nucleic Acids Res.* 38:D46-D51.

Bentley, S. D., Corton, C., Brown, S. E., Barron, A., Clark, L., Doggett, J., Harris, B., Ormond, D., Quail, M. A., May, G., Francis, D., Knudson, D., Parkhill, J. and Ishimaru, C. A. 2008. Genome of the actinomycete plant pathogen *Clavibacter michiganensis* subsp. *sepedonicus* suggests recent niche adaptation. *J. Bacteriol.* 190:2150-2160.

Campbell, C. L. and Madden, L. V. 1990. Introduction to Plant Disease Epidemiology. John Wiley and Sons, New York, USA. 532 pp.

Chang, R. J., Ries, S. M. and Pataky, J. K. 1991. Dissemination of *Clavibacter michiganensis* subsp. *michiganensis* by practices used to produce tomato transplants. *Phytopahology* 81:1276-1281.

Chen, T., Madey, J. M. J., Price, F. M., Sharma, S. K. and Lienert, B. 2007. Remote Raman spectra of benzene obtained from 217 meters using a single 532 nm laser pulse. *Appl. Spectrosc.* 61:624-629.

Colvine, S. and Branthôme, F. X. 2016. The tomato: a seasoned traveler. In: *The Tomate Genome*, eds. by M. Causse, J. Giovannoni, M. Bouzajen and M. Zouine, pp. 1-6. Springer-Verlag, Berlin, Germany.

Eichenlaub, R. and Gartemann, K. H. 2011. The *Clavibacter michiganensis* subspecies: molecular investigation of gram-positive bacterial plant pathogens. *Annu. Rev. Phytopathol.*

49:445-464.

EPPO, 2016. PM 7/42 (3) *Clavibacter michiganensis* subsp. *michiganensis*. *EPPO Bulletin* 46:202-225.

FAO, 2017. Food and Agriculture Organization. Deposited on: URL http://www.fao.org/faostat/es/#rankings/countries_by_commodity/ [15 July 2017].

Fletcher, J., Bender, C., Budowle, B., Cobb, W. T., Gold, S. E., Ishimaru, C. A., Luster, D., Melcher, U., Murch, R., Scherm, H., Seem, R. C., Sherwood, J. L., Sobral, B. W. and Tolin, S. A. 2006. Plant pathogen forensics: capabilities, needs and recommendations. *Microbiol. Mol. Biol. Rev.* 70:450-471.

Gan, Q., Wang, X., Wang, Y., Xie, Z., Tian, Y. and Lu, Y. 2017. Culture-free detection of crop pathogen at the single-cell level by Micro-Raman spectroscopy. *Adv. Sci.* 4:1700127.

Gelder, J. D., Gussem, K. D., Vandenabeele, P., Vancanneyt, M., Vos, P. D. and Moens, L. 2007. Methods for extracting biochemical information from bacterial Raman spectra: focus on a group of structurally similar biomolecules-Fatty acids. *Anal. Chim. Acta* 603:167-175.

Guillén-Sánchez, D., Téliz-Ortiz, D., Mora-Aguilera, G., Mora-Aguilera, A., Sánchez-García, P. and González-Hernández, V. 2003. Desarrollo temporal de epidemias de cenicilla (*Oidium mangiferae* Berthet) en huertos de mango (*Mangifera indica* L.) en Michoacán, México. *Rev. Mex. Fitopatol.* 21:181-188 (in Spanish).

Hartigan, J. A. and Wong, M. A. 1979. Algorithm as 136: A k-means clustering algorithm. *J. Royal Stat. Soc.* 28:100-108.

Izenman, A. J. 2008. Linear Discriminant Analysis. In: *Modern Multivariate Statistical Techniques*, pp. 237-280. Springer, New York, USA.

Jasso, C. C., Martínez, G. M. A., Chávez, V. J. R., Ramírez, T. J. A. and Garza, U. E. 2012. *Guía para cultivar jitomate en condiciones de malla sombra en San Luis Potosí.* URL http://www.inifapcirne.gob.mx/Biblioteca/Publicaciones/905.pdf/ (in Spanish)

Jehlička, J. and Oren, A. 2013. Raman spectroscopy in halophile research. *Front. Microbiol.* 4:380.

Jolliffe, I. 2002. Principal components as a small number of interpretable variables: some examples. In: *Principal Component Analysis*. ed by J. Jolliffe, pp. 63-77. Springer, New York, USA.

Kawaguchi, A., Tanina, K. and Inoue, K. 2010. Molecular typing and spread of *Clavibacter michiganensis* subsp. *michiganensis* in greenhouses in Japan. *Plant Pathol.* 59:76-83.

Knapp, S. and Peralta, I. E. 2016. The tomato (*Solanum lycopersicum* L., Solanaceae) and its botanical relatives. In: *The Tomate Genome*, eds. by M. Causse, J. Giovannoni, M. Bouzajen and M. Zouine, pp. 7-21. Springer-Verlag, Berlin, Germany.

Lorenz, B., Wichmann, C., Stöckel, S., Rösch, P. and Popp, J. 2017. Cultivation-free Raman spectroscopy investigations of bacteria. *Trends Microbiol.* 25:413-424.

Maiti, N., Kapoor, S. and Mukherjee, T. 2013. Surface-enhanced Raman Scattering (SERS) spectroscopy for trace level detection of chlorogenic acid. *Adv. Mater. Lett.* 4:502-506.

Malard, L. M., Pimenta, M. A., Dresselhaus, G. and Dresselhaus, M. S. 2009. Raman spectroscopy in grapheme. *Phys. Rep*. 473:51-87.

Martínez-Castro, E., Jarquín-Galvez, R., Alpuche-Solís, Á. G., Vallejo-Pérez, M. R., Colli-Mull, J. G. and Lara-Ávila, J. P. 2018. Bacterial wilt and canker of tomato: fundamentals of a complex biological system. *Euphytica* 214:72.

Mirski, T., Bartoszcze, M., Bielawska-Drózd, A., Cieślik, P., Michalski, A. J., Niemcewicz, M., Kocik, J. and Chomiczewski, K. 2014. Review of methods used for identification of biothreat agents in environmental protection and human health aspects. *Ann. Agric. Environ. Med.* 21:224-234.

Misra, A. K., Sharma, S. K., Kamemoto, L., Zinin, P. V., Yu, Q., Hu, N. and Melnick, L. 2009. Novel micro-cavity substrates for improving the Raman signal from submicrometer size materials. *Appl. Spectrosc.* 63:373-377.

Monciardini, P., Sosio, M., Cavaletti, L., Chiocchini, C. and Donadio, S. 2002. New PCR primers for the selective amplification of 16S rDNA from different groups of actinomycetes. *FEMS Microbiol. Ecol.* 42:419-429.

Movileanu, L., Benevides, J. M. and Thomas, G. J. 1999. Temperature dependence of the Raman spectrum of DNA. Part I–Raman signatures of premelting and melting transitions of poly(dA–dT)·poly(dA–dT). *J. Raman Spectrosc.* 30:637-649.

Paret, M. L., Sharma, S. K., Green, L. M. and Alvarez, A. M. 2010. Biochemical Characterization of Gram-Positive and Gram-Negative Plant-Associated Bacteria with Micro-Raman Spectroscopy. *Appl. Spectrosc.* 64:433-441.

Paret, M. L., Sharma, S. K. and Alvarez, A. M. 2012. Characterization of biofumigated *Ralstonia solanacearum* cells using Micro-Raman spectroscopy and electron microscopy. *Phytopathology* 102:105-113.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E. 2011. Scikit-learn:machine learning in Python. *J. Mach. Learn. Res.* 12:2825-2830.

Pérez, M. R. V., Mendoza, M. G., Elías, M. G., González, F. J., Contreras, H. R. and Servín, C. C. 2016. Raman spectroscopy an option for the early detection of citrus Huanglongbing. *Appl. Spectrosc.* 70:829-839.

Polisetti, S., Bible, A. N., Morrell-Falvey, J. L. and Bohn, P. W. 2016. Raman chemical imaging of the rhizosphere bacterium *Pantoea* sp. YR343 and its co-culture with *Arabidopsis thaliana*. *Analyst* 141:2175-2182.

Pruesse, E., Peplies, J. and Glöckner, F. O. 2012. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* 28:1823-1829.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. and Glöckner, F. O. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41:D590-D596.

Sambrook, J. and Russell, D. 2001. Molecular Cloning: A Labo-

ratory Manual. 3rd ed. Cold Spring Harbor Laboratory Press, NY, USA. 2344 pp.

Santos, M. C., Morais, C. L., Nascimento, Y. M., Araujo, J. M. and Lima, K. M. 2017. Spectroscopy with computational analysis in virological studies: A decade (2006-2016). *Trends Analyt. Chem.* 97:244-256.

Saperstein, S., Starr, M. P. and Filfus, J. A. 1954. Alterations in carotenoid synthesis accompanying mutation in Corynebacterium michiganense. *J. Gen. Microbiol.* 10:85-92.

Sen, Y., Feng, Z., Vandenbroucke, H., van der Wolf, J., Visser, R. G. F. and van Heusden, A. W. 2013. Screening for new sources of resistance to *Clavibacter michiganensis* subsp. *michiganensis* (Cmm) in tomato. *Euphytica* 190:309-317.

Sen, Y., van der Wolf, J., Visser, R. G. F. and van Heusden, S. 2015. Bacterial canker of tomato: current knowledge of detection, management, resistance and interactions. *Plant Dis.* 99:4-13.

Sharabani, G., Shtienberg, D., Borenstein, M., Shulhani, R., Lofthouse, M., Sofer, M., Chalupowicz, L., Barel, V. and Manulis-Sasson, S. 2013. Effects of plant age on disease development and virulence of *Clavibacter michiganensis* subsp. *michiganensis* on tomato. *Plant Pathol.* 62:1114-1122.

Siqueira, L. F. S. and Lima, K. M. G. 2016. MIR-biospectroscopy coupled with chemometrics in cancer studies. *Analyst* 141:4833-4847.

Velez, D. R., White, B. C., Motsinger, A. A., Bush, W. S., Ritchie, M. D., Williams, S. M. and Moore, J. H. 2007. A balanced accuracy function for epistasis modeling in imbalanced datasets using multifactor dimensionality reduction. *Genet. Epidemiol.* 31:306-315.

Wang, Y., Huang, W. E., Cui, L. and Wagner, M. 2016. Single cell stable isotope probing in microbiology using Raman microspectroscopy. *Curr. Opin. Biotechnol.* 41:34-42.

Wold, S., Esbensen, K. and Geladi, P. 1987. Principal component analysis. *Chemometr. Intell. Lab*. 2:37-52.

Xanthopoulos, P., Pardalos, P. M. and Trafalis, T. B. 2013. Linear Discriminant Analysis. In: *Robust Data Mining*, pp. 27-33. SpringerBriefs in Optimization. Springer, New York, USA.

Xu, X., Miller, S. A., Baysal-Gurel, F., Gartemann, K. H., Eichenlaub, R. and Rajashekara, G. 2010. Bioluminescence imaging of *Clavibacter michiganensis* subsp. *michiganensis* infection of tomato seeds and plants. *Appl. Environ. Microbiol*. 76:3978-3988.

Zhao, J., Lui, H., McLean, D. I. and Zeng, H. 2007. Automated autofluorescence background subtraction algorithm for biomedical spectroscopy. *Appl. Spectrosc.* 61:1225-1232.