Investigation of the stages of citrus greening disease using micro synchrotron radiation X-ray fluorescence in association with chemometric tools†

Fabiola Manhas Verbi Pereira* and Débora Marcondes Bastos Pereira Milori

Received 7th October 2009, Accepted 18th November 2009
First published as an Advance Article on the web 23rd November 2009
DOI: 10.1039/b920980h

Citrus greening has been considered the worst citrus pest since 2004, both in Brazil and in the United States of America. This disease has had serious consequences in agricultural production and highlights the urgent need for early diagnostic methods. This analytical study investigated the mineral constituents of healthy leaves and leaves infected with citrus greening (or Huanglongbing) using micro synchrotron radiation X-ray fluorescence (µSR-XRF) scanning combined with chemometric tools. The information obtained using the µSR-XRF spectra profiles and the chemometric tools allowed the construction of predictive models to identify healthy and infected trees with and without symptoms. The signals for K, Ca, Fe, Cu and Zn and the region of coherent and incoherent scatterings were relevant to the differentiation of the healthy and infected samples. The calibration model correctly classified up to 95–98% of samples, and the validation data correctly assigned 90% of samples with a 95% significance level.

Introduction

The first evidence of citrus greening was reported in southern China in 1919, when it was known as Citrus Huanglongbing (HBL). Since then, this pest has caused damage for producers of sweet oranges and other citrus fruits around the world. In the United States of America and Brazil, citrus greening has been considered the worst pest of citrus plantations since 2004, causing serious damage in agricultural production. The mechanism of the infection process is the establishment of the bacteria Candidatus Liberibacter spp. in the phloem of the plant. The phloem is part of the vascular system of higher plants, and obstruction of the phloem can affect the flow of sap to the roots, stem and leaves, the movement of which is critical for the health of plants.

The natural insect vector of the pathogen has been identified as Diaphorina citri, which is often found in Brazilian orchards. There are three different species of the pathogen (Candidatus Liberibacter) that can be spread: americanus, africanus and asiaticus. In addition to insect-borne transmission, citrus greening can also be transmitted by grafting from an infected citrus plant.

Plants infected by citrus greening have two different stages: asymptomatic and symptomatic. The asymptomatic stage is very dangerous because the plant is contaminated. Depending on the incubation period, this condition is not perceptible by visual inspection. The main symptoms of citrus greening are yellowing of the leaves and the production of small and deformed fruits. Generally, polymerase chain reactions (PCRs) are used to confirm infection of citrus crops after citrus greening disease has been suspected.

To detect this disease, efforts such as the determination of micronutrients in plant extracts contaminated with citrus greening have presented evidence of the variation of Fe and Mn concentrations in healthy and infected plants. Zinc deficiency was also associated with infection. The main problem with this severe pest is the ineffectiveness of pesticides against the vector. Another factor is that the asymptomatic period can last from 6 to 36 months. Up to now, the eradication of infected trees is the only procedure that can prevent the spread of the infection. The consensus about the control of the disease is the development of early diagnostic methods.

Experimental

Plant material

A total of 180 leaves from 60 Citrus sinensis (L.) Osbeck, variety Valencia/Swingle, trees were collected. The leaves were divided...
into three categories of equal size: healthy, asymptomatic and symptomatic. In the latter two categories, the plants were infected with the causative agent of citrus greening. The citrus crop was cultivated at Matão (São Paulo State, Brazil) starting in February 2006. The parameters such as weather, soil condition, watering and fertilization were adequate and favourable to growing the plants. The leaves were taken six days before the beginning of the laboratory experiments. Each leaf was cleaned with the aid of a piece of cotton wetted with deionised water and was then dried in an air atmosphere. The samples were stored at 4 °C inside plastic boxes covered with dark bags. The presence of citrus greening disease was confirmed by PCR amplification of Candidatus Liberibacter asiaticus sequences.

Experiments using µSR-XRF

The measurements were performed with the D09B X-ray fluorescence beam line at the Brazilian Synchrotron Light Source (LNLS, Campinas, São Paulo State). Each raw leaf was fixed in a sampler with adhesive tape (Adelbras, Vinhedo, São Paulo State, Brazil). The setup configuration used a monochromatic beam with a 200 × 200 μm cross section. The multilayer monochromator consisted of 75 d-periods of W/C layers. The maximum photon energy was 12 keV, and the tests were executed in an air atmosphere. The active total area of the detector was 30 mm². The 4 mm circular collimator was used in the detector to improve the signal-to-noise ratio of the XRF spectra and to avoid decreasing of resolution at higher count rates. The distances from the sample to the source and to the detector were 120 mm and 20 mm, respectively. An Si(Li) detector with an 8 μm Be window was used; this detector has an energy resolution of 165 eV for the Mn Kz line. The samples were positioned at an angle of 45° with respect to both the incident beam and the detector. The measurement positions on the leaves were marked with computer-controlled X, Y, Z, θz stages and an optical microscope. The scanning was divided into three regions: central nervure and two lateral sides. Nine spectra were obtained per leaf, as shown in Fig. 1. The vertical and horizontal distances between points were 1 mm. The irradiation time per point was 100 s and the dead time was approximately 2%.

Data set treatment

The areas of the µSR-XRF spectra peaks, including the scattering region (coherent and incoherent effects), were estimated using the AXIL software provided by International Atomic Energy Agency (IAEA). The software Pirouette 4.0 rev.2 (Infometrix, Inc.; Bothell, Washington State, USA) was used for data set treatment and for chemometric applications. The applied chemometric tools were principal component analysis (PCA), soft independent modelling of class analogy (SIMCA), Kth nearest neighbour (KNN) and partial least squares discriminant analysis (PLS-DA). The algorithm used in all chemometric analyses, except KNN, was based on Nonlinear Iterative Partial Least Squares (NIPALS). The KNN is based on Euclidian distance among the samples.

Results and discussion

X-ray spectra analysis of citrus leaves

Leaves were chosen for the experiments because the main symptoms of citrus greening are very apparent in this part of the plants. During the measurements, several characteristic X-ray fluorescence lines were verified in all three categories of citrus leaves (healthy, asymptomatic and symptomatic). The spectra profiles were very similar for all these categories of samples, and only one for asymptomatic condition was shown in Fig. 2. From this same figure, the lines were as follows: S Kα 2.308 keV, K Kα 3.313 keV, Ca Kα 3.691 and Kβ 4.013 keV, Mn Kα 5.895 keV, Fe Kα 6.400 and Kβ 7.058 keV, Cu Kα 8.042 keV and Zn Kα 8.632 and Kβ 9.572 keV, respectively. The region of coherent and incoherent scatterings was located approximately between 11 and 13 keV. The line Kα 2.957 keV was corresponding to Ar, which was utilised for the ionisation camera. Titanium (Kα 4.509 and Kβ 4.932 keV), was present in the adhesive tape used to fix the leaf on the sampler. The mean signal was approximately 7 counts for this element.

The following micronutrients were constituents of the plants: Mn, Fe, Cu and Zn. Macronutrient peaks were also visible: S, K and Ca. The mean intensity values for S, Mn and Cu were lower than 20 counts. The other elements, Fe and Zn, showed values higher than 20 counts, and the counts for Fe were twice as high as the counts for Zn. The highest values of the mean intensities ranged from 600 to 1650 counts for K and Ca, respectively, and from 60 to 75 counts for the scattering region.

Fig. 1 Positions scanned on the citrus leaves during the µSR-XRF experiments.

Fig. 2 Example of the spectrum profile obtained for asymptomatic leaf condition.
Multivariate spectra analysis of citrus leaves

Exploratory analysis of the data using principal component analysis. A total of 1645 spectra were recorded: 540 for healthy leaves, 565 for asymptomatic leaves and 540 for symptomatic leaves. The 25 excess spectra for asymptomatic leaves were additional readings of the same leaves, using the same positions. Considering the large amount of data and the possible tenuous variations therein, chemometric tools were used with the intention of avoiding subjectivity in the data interpretation or even misinterpretation during the data analysis.

The categories of samples were divided into three groups according to the condition of the leaves as described previously. To obtain information about the elemental distribution within the leaves, the data were also analysed based on the scanned region. Three different regions were scanned: side 1, 1 mm away from the central nervure (shown in Fig. 1, points 1, 4 and 7); the middle of the central nervure (points 2, 5 and 8); and side 2 (points 3, 6 and 9), also 1 mm away from the central nervure. The scanning was executed at these positions considering that the intense alterations of the sap flow caused by the bacteria can be better detected in this region. In addition, other criteria for the composition of the categories were also taken into account, such as noting leaf sample contribution and using different trees as leaf sources. The variables responsible for the possible discrimination among these categories were evaluated using loading plots.

The data analysis strategy was to perform spectra pre-treatment. This pre-treatment was executed using the arithmetic mean of the spectra samples. The detailed procedure of the calculations was as follows: ten-by-ten spectra were separated for each of the measured points, from 1 to 9. In other words, for each category, 54 mean spectra were resultant from 6 mean spectra of each scanned point, thus generating a matrix with 162 rows (samples) by 2048 columns (variables).

The data plots after the mathematical pre-treatment and PCA analysis of the variables are shown in Fig. 3 with scores representing samples and loadings representing variables. The spectra set were mean-centered. Clusters representing the three different sample categories were apparent, but the asymptomatic samples (circles) tended to mix with the healthy and symptomatic samples. In general, the cluster of symptomatic samples (noted by stars in Fig. 3a) exhibited more variation than the other clusters. The contribution of the variables can be visualised on the loading graph (Fig. 3b). PC1 and PC2 together explained 96% of the observed variance. Potassium and calcium showed the higher values of loadings on PC1, as shown by the solid line. With PC2 (dotted line), the increase of the scattering region values was verified.

The calculated areas of S, K, Ca, Mn, Fe, Cu and Zn and of the scattering region were also utilised in PCA evaluations and resulted in the same above-mentioned information. The advantage of using the peak areas is the speed of the data analysis because there were 8 variables for each sample, in comparison to using the whole spectra (2048 variables). However, the use of the spectra profile for the following models contributed to full information compared to the calculated values of area that can be justified based on the fact that the values of area were estimated.

Construction of calibration models. To evaluate the potential of the µSR-XRF spectra in the identification of trees affected by citrus greening, a calibration model was constructed using soft independent modelling of class analogy (SIMCA) with the whole spectral region and the matrix with the mean data (162 samples and 2048 variables). The best pre-processing of the variables was mean-centered for all models. From the model constructed with whole spectra, it was possible to verify the most important variables using the information from the loadings, discriminating power and total modelling power plots.

Other SIMCA models were tested using only the Ca line, and correct predictions of the condition of the plants were up to 90% and 50% of the calibration and the validation data sets, respectively.

Another model using only the maximum value from the Ca line to construct a univariate calibration curve was proposed. This model was performed based on differences of the mean intensity values verified for Ca, considering each class, as follows: 1580 counts for healthy leaves, 1650 for asymptomatic leaves and 1219 for symptomatic leaves. Therefore, good regression coefficient of linearity and predictions were not obtained.

The K signal showed a tendency to increase for infected samples, with healthy leaves having 485 counts and asymptomatic and symptomatic leaves having 600 counts, but the model using only the line of this element was not able to distinguish between healthy and infected leaves.

Other models were executed using several combinations, for example only the fluorescence lines for micronutrients (Mn, Fe, Cu, Zn and Fe) were used.
Cu and Zn) or only macronutrients lines (K and Ca). Models were constructed both including and excluding the coherent and incoherent region. These models were not successful.

Furthermore, other chemometric tools were applied, such as Kth Nearest Neighbour (KNN) and partial least squares discriminant analysis (PLS-DA). However, the best fit models were obtained by SIMCA.

The KNN and PLS-DA did not work well, in the case of this data, because some asymptomatic samples resembled healthy or infected samples, as showed by PCA (Fig. 3a). The KNN is based on Euclidean distance, and therefore the magnitude of the distance values between asymptomatic and healthy or asymptomatic and infected samples was not effective for successful classification of the samples. PLS-DA uses pre-defined categories as SIMCA, but the algorithms are based on PLS and PCA, respectively. On the case of PLS-DA the parameter of classification is the values designated for each class. On SIMCA, the advantage is the algorithm executes the construction of individual models for each class. And the criterion of classification is to belong to the same population as the calibration (or training) set. Then, the particular characteristics of healthy, asymptomatic and symptomatic were better performed in SIMCA than the others.

Then, the final model was performed using 677 selected variables (ranging from 2.69 to 4.33 keV; from 4.36 to 4.40 keV; from 6.14 to 9.22 keV; from 10.75 to 12.65 keV). These variables were chosen with the aid of the loading, discriminant power and modelling power plots from the previous model using the whole spectra. The following fluorescence lines were made part of this range: K, Ca, Fe, Cu and Zn. This range also included the region of coherent and incoherent scatterings for the construction of this model. Using these variables, it was possible to extract the maximum amount of information from the spectra of the analysed leaves because the most relevant variables for this model were defined. The selected samples for calibration and validation were 44 (total of 132) and 10 (total of 30) for each condition, respectively. The selection of these samples was done on the aleatory way, avoiding samples located at the edge of each cluster of class.

The scattering region was very important for this model because the sap of the phloem is composed of a combination of organic compounds, such as sugars (over 90% carbohydrates, including sucrose and other oligosaccharides), and inorganic constituents, such as mineral elements and nutrients. Other parts of the plants are approximately 98% C, N, H and O. These elements have low absorption coefficients for X-ray excitation so they scatter the radiation. The presence of them can influence the absorption of other elements, such as S, K, Ca and Fe, when these other elements are present in low concentrations. The inclusion of the coherent and incoherent scatterings can attenuate the effect of the organic elements and compensate for thickness differences among the leaves, and consequently can provide useful information for the model.

The optimal number of principal components for each class was 4 (from PC1 to PC4). The total explained variance was 99% for healthy and asymptomatic leaves but was 100% for

Fig. 4 Loading plots obtained from the SIMCA model optimised with 4 principal components (PC) for the three leaf conditions: (a) PC1, (b) PC2, (c) PC3 and (d) PC4.

| Table 1 Results of the calibration and validation data sets using SIMCA |
|-----------------------------------------------|---------------|---------------|-----------------|-----------------|---------------|---------------|---------------|
| Calibration data set (132 samples)            | Validation data set (30 samples)               |
| Condition | Healthy | Asymptomatic | Symptomatic | Correct predictions (%) | Condition | Healthy | Asymptomatic | Symptomatic | Correct predictions (%) |
| Healthy (44) | 42     | 2a          | 0           | 95                | Healthy (10) | 9       | 1a          | 0           | 90                |
| Asymptomatic (44) | 1a     | 43         | 0           | 98                | Asymptomatic (10) | 1a     | 9          | 0           | 90                |
| Symptomatic (44) | 0     | 1a         | 43          | 98                | Symptomatic (10) | 1a     | 0          | 9           | 90                |

a incorrect predictions.
symptomatic leaves. The values of the explained variance are in parenthesis in Fig. 4. The pre-processing local scope was used, that is done on each category individually. For the following discussion, the plots of each principal component are on the same scale.

From the analysis of the loading plots for principal component 1 (PC1) for each leaf category, which explained 83% to 84% of the variance, it was possible to verify that Ca was correlated with the three conditions of the leaves with negative values of loadings, as shown in Fig. 4a (see the dotted rectangle). Potassium had negative values for healthy and symptomatic leaves, with low values of loadings for the latter samples. The contribution of K was not detected for the asymptomatic samples in this PC.

In the PC2 evaluation in Fig. 4b, the values of loadings for K were negative only for healthy samples and were positive for asymptomatic and symptomatic samples. Calcium line was present only for healthy samples; the values were very low or negative for the other conditions. The values of loadings for the fluorescence lines of Fe, Cu and Zn and coherent and incoherent scatterings showed the importance of these variables to the modelling of healthy class. For the two other conditions, the values of loadings for Fe, Cu, Zn and coherent and incoherent scatterings were negative.

It was verified that, for healthy leaves, the same lines that contributed to PC2 also contributed to PC3 (Fig. 4c), but the values for K were positive and those for Ca were negative, contrary to asymptomatic and symptomatic leaves.

PC4 showed positive values of loadings for Cu and Zn for healthy samples (Fig. 4d). The Fe line had negative values, and the other lines were noisy. For asymptomatic samples, the signals for Fe, Cu and Zn had negative values, and the scattering region showed positive values. All other lines were noisy. The symptomatic samples showed positive values for Fe, Cu and Zn. Contrary to healthy samples, the positive values of loadings decreased for Cu and increased for Zn.

For the calibration data set, correct category predictions were made for 95% of healthy leaves and 98% of asymptomatic and symptomatic leaves. Two healthy leaves were misclassified as asymptomatic samples. In the case of the asymptomatic leaves, one leaf was incorrectly classified as healthy. For symptomatic leaves, one leaf was classified as asymptomatic. These results are displayed in Table 1. The validation data set had 90% of the correct values for the three conditions. All of the samples were modelled and matched with at least one category. These results were significant at the 95% confidence level.

Conclusions
The evaluation of the mineral composition using μSR-XRF and chemometric tools was suitable for understanding the effects of the citrus greening disease. The elements K, Ca, Fe, Cu and Zn and the region of coherent and incoherent scatterings were very important for the differentiation of these three conditions.

In addition, the model presented herein opens the possibility of predicting the condition of the infection stage using mineral composition. The model reinforces the evidence that plants infected by citrus greening (both asymptomatic and symptomatic) undergo alterations in their micro- and macro-nutrient compositions. This information deserves attention because it is related to biochemical and physiological processes related to the infection. The monitoring of the mineral compositions of the plants was qualitative but could generate responses about the infected samples, mainly those in the asymptomatic stage.

Acknowledgements
We are grateful to Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) for the fellowship awarded to F. M. V. Pereira (Process 2007/08618-2), to Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (Process 578576/2008-2), to Grupo Fischer - Divisão Citrosuco for providing the plant material and to the D09B X-ray fluorescence beam line at the Brazilian Synchrotron Light Source - LNLS (Process D09B - XRF-8491/09) for the μSR-XRF measurements.

References