**Attenuated Total Reflection Fourier-Transform Infrared Spectroscopy Coupled with Chemometrics Directly Detects Pre- and Post-Symptomatic Changes in Tomato Plants Infected with *Botrytis cinerea***

Paul Skolik1, Camilo L.M. Morais2,Francis L. Martin3, Martin R. McAinsh\*,1

*1Lancaster Environment Centre, Lancaster University, Bailrigg, Lancaster, LA1 4YQ,*

*United Kingdom*

*2School of Pharmacy and Biomedical Sciences, University of Central Lancashire, Preston, PR1 2HE United Kingdom*

*3Biocel Ltd, Hull HU10 7TS, UK*

*\* Corresponding author: m.mcainsh@lancaster.ac.uk*

**Funding Details**

This work was supported by the Agriculture and Horticulture Development Board (AHDB) (Grant number CP 119): Sensor-based pre-symptomatic detection of pests and pathogens for precision scheduling of crop protection products (studentship). Approaches and methodologies used in this study were developed in work funded by the Biotechnology and Biological Sciences Research Council (grant number BB/D010055/1) and the Engineering and Physical Sciences Research Council (EPSRC; Grant Nos: GR/S75918/01 and EP/K023349/1).

**Conflict of Interest Statement**

Both FLM and CLMM are shareholders in Biocel UK Ltd. FLM holds a research and development position within Biocel UK Ltd.

**Data Availability**

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request**.**

**Attenuated Total Reflection Fourier-Transform Infrared Spectroscopy Coupled with Chemometrics Directly Detects Pre- and Post-Symptomatic Changes in Tomato Plants Infected with *Botrytis cinerea***

Abstract

Sensor-based detection of pests and pathogens in a high throughput and non-destructive manner is essential for mitigating crop loss. Infrared (IR) sensors in the form of vibrational spectroscopy provide both biochemical information about disease, as well as a large number of variables for chemometrics. This approach is highly adaptable to most biological systems including interactions between plants and their environments. Fast-acting necrotrophic fungal pathogens present a specific group of pests with adverse effects on food production and supply and are therefore pertinent to food security. *Botrytis cinerea* and *Solanum lycopersicum* are models for the study of fungal and crop biology respectively. Herein we use a compact mid-IR spectrometer with attenuated total reflection (ATR) attachment to measure the plant-microbe interaction between *S. lycopersicum* and *B. cinerea* on leaves, *in vivo* of intact plants. Chemometric models including exploratory principal component analysis (PCA) solely, and as a classifier in combination with linear discriminant analysis (PCA-LDA) are applied. Fingerprint spectra (1800-900 cm-1) were excellent discriminators of plant disease in both visually symptomatic as well pre-symptomatic plants. Major biochemical alterations in leaf tissue as a result of infection are discussed. Diagnostic potential for automatic decision-making platforms is shown by high accuracy rates of 100% for detecting plant disease at various stages of progression.

Keywords: Infrared spectroscopy, tomato, *Botrytis cinerea*, chemometrics, crop biology, pest detection, sensors

Introduction

The challenges to sustainable crop production and supply are considerable and include population rise, depletion of natural resources, and climate change. Combined with these, crop loss and food waste to either pests and pathogens, or due to general waste, puts a heavily burden on humans and the environment (Wunderlich and Martinez 2018). Conventional but outdated farming strategies are compounding these issues because traditional methods are only being slowly replaced by more sustainable approaches. The example of fungicides use illustrates that these are only marginally effective, expensive, and yet routinely overused (Bourguet and Guillemaud 2016; Vasileiadis 2017). Alternatives to chemical pest control are thus favorable, especially with the growing concern about harmful residues of crop protection products in food (Singbo et al. 2015). One approach therefore may be to focus on the early detection of plant disease to remediate imminent threats, or prevent further disease spread.

Crop loss to destructive pests and pathogens poses a substantial threat to food security and the economy. Around one third of global annual crop loss can be attributed to pests and pathogens (Oerke 2006) including macroscopic pests such as weeds, herbivorous animals, and insects, and microscopic pathogens such as viruses, bacteria, and fungi. In contrast to macroscopic pests, microscopic pathogens are invisible without the use of tools, becoming apparent in horticultural environments only through adverse effects (symptoms) they cause. Moreover, the current horticultural standard for detection of these pathogens relies on the appearance of plant disease symptoms and is thus subjective (Mahlein 2016). Microscopic pests thus remain especially challenging for plant and crop scientists to manage (Williamson et al. 2007), primarily due to their microscopic nature, but also because they employ a vast combination of pathogenic strategies for survival (El Oirdi et al. 2011). The ubiquitous distribution of microbial pathogens and their long-standing interactions with host plants throughout plant evolution makes these plant-pathogen interactions highly complex and diversified (Dayan et al. 2009).

Fungi are pests that have evolved to invade plants by various means and with varying consequences. Three main strategies distinguish fungal pathogens. These include biotrophism, hemi-biotrophism, and necrotrophism. Biotrophs complete their life cycle without killing the host plant and rather establish a nutritional mode that depends on living plant cells (Vleeshouwers and Oliver 2014). Hemibiotrophs begin their life cycle with a biotrophic strategy and can switch to a necrotrophic mechanism, which results in cell death (Vleeshouwers and Oliver 2014). Necrotrophic pathogens therefore have the potential to destroy entire plants and crops rapidly. *Botrytis cinerea* is an important necrotroph capable of causing extensive damage to food crops, both pre- and post-harvest, and is therefore widely studied as a model necrotoph. High genetic diversity, various modes of attack, a broad host range, and the ability to remain quiescent, make *B. cinerea* a formidable threat to crops infecting most plant tissues including seeds, seedlings, roots, stems, leaves, flowers, young and mature plants, fruit, and plant waste (Williamson et al. 2007). The range of host species numbering over 200 provide a diverse source of potential inoculum which can be ubiquitous in horticultural environments as mycelia, conidia, or prolonged periods as sclerotia, making control of this pathogen particularly difficult (Elad and Stewart 2007). Many fungicides have become ineffective over time due to pathogen adaptation leading to potential overuse of these pesticides making this approach costly and unsustainable in a horticultural context (Lechenet et al. 2014, 2017). Therefore, new approaches for the management of pathogenic fungi are required to optimize crop protection measures. To this end, a more detailed understanding of plant-pathogen interactions and the way they manifest during crop cultivation is an essential prerequisite for improving crop utilization and minimizing crop loss due to microscopic pathogens.

The rapid and early detection of fungal pathogens in the crop-growing environment remains a key challenge. This is especially so when there is requirement for the non-destructive analysis of crops during growth, leaving them unaltered in the process, for example, when monitoring intact plants and crops, or produce within the food supply chain. These criteria favor adaptable sensors capable of analysis at both pre- and post-harvest stages without damaging the crop; this is in marked contrast to destructive methods which require reagent preparation, labeling, extensive sample preparation, and provide potentially limited information with regards to pathogen detection. Within the array of non-destructive sensors, spectroscopic methods have been used successfully in horticultural settings and are becoming increasingly adaptable for intact plant analysis (Egging et al. 2018; Farber and Kurouski 2018; Trebolazabala et al., 2013, 2017; Yeturu et al. 2016). These methods have the advantage of being capable of the fast, non-destructive analysis of a diverse set of sample types, which provide the biochemical information on which to build models of disease specificity (Martin et al. 2010). Sensor-based early disease detection has the potential to contribute to the sustainable intensification of crop production, by improving the effectiveness with which crop protection measures are applied and thereby considerably reducing disease spread through precision agriculture. Progress in sensor technology, computational analysis, and established machine learning methods offers new opportunities for this approach through the development of mobile devices for the early detection and identification of plant disease (Behmann et al. 2014). Surface techniques have been widely used for analytical chemistry and have recently been developed as sensors for heterogeneous biological materials including cells, tissues, and plants. Analytical surface techniques include infrared vibrational spectroscopy, a non-destructive sensor conferring data with high chemical specificity in the form of a highly characteristic infrared spectrum. The ability to measure almost any sample type without preparation or labeling using mid-infrared (MIR) vibrational spectroscopy in biology (biospectroscopy), has made it a valuable tool for studying biological systems at the tissue, cell, and molecular levels in plants and crops (Skolik et al. 2018). One method that has stood out in the biological context and adaptability to *in vivo* measurements non-destructively is attenuated total reflectance (ATR) Fourier transform infrared (FTIR) spectroscopy, which uses a high refractive index material interface to make direct sample contact, resulting in highly reproducible spectra from living tissues (Chan and Kazarian 2016). Further benefits of ATR-FTIR is the very well-defined depth of interrogation into the sample and its ability to circumvent the effects of water in plant leaves when practically applied (Butler et al. 2015). This makes MIR spectroscopy in ATR mode favorable for whole tissue analysis including intact leaves and fruit (Butler et al 2015; Skolik et al 2019).

Extracting information from complex biological samples requires computational analysis such as linear and non-linear data models (chemometrics) (Morais and Lima 2017). Chemometrics such as multivariate analysis using supervised and unsupervised methods including principal component and linear discriminant analysis (PCA and LDA respectively) as well more complex models including support vector machine (SVM) and artificial neural networks (ANN) have also been used in various amalgamations to answer questions in the biomedical, environmental, plant and crop sciences (Trevisan et al. 2012). Biospectroscopy in combination with advanced computational analysis therefore has the potential to contribute to the development of sustainable farming practices and precision agriculture, and to help achieve the goal of producing more food with the same amount of land by reducing crop losses to pests and pathogens through the adoption of ‘high-tech farming’ approaches.

Tomato *S. lycopersicum* is an ancient crop which is widely used for the study of crop biology (Kimura and Sinha 2008). Plants and fruit of tomato are among the many species susceptible to pathogen infection by *B. cinerea* at all stages of development (Elad and Stewart 2007; Williamson et al. 2007). Tomato interaction with fungi thereby presents an excellent model to investigate challenging plant-pathogen interactions with rapid disease progression for the purposes of sensor development. This study applies MIR biospectroscopy to a glasshouse growing environment, to investigate the effectiveness of this approach within an agriculture context using the interaction between glasshouse-grown tomato (*S. lycopersicum* cv. Moneymaker) and *B. cinerea*. Chemometric approaches including PCA and PCA-LDA were used to investigate spectral alterations in response to *B. cinerea* within intact plant leaves. PCA is an unsupervised multivariate analysis approach that investigates dataset variance without considering class label and thus provides insight into dataset variance, specifically whether the biological effects under investigation are the main source of variance, or whether supervised models (LDA) are required to extract class-specific differences which are subtler than can be extracted via PCA alone. Further, both PCA and LDA are linear multivariate models that provide loadings, which may be interpreted as spectral biomarkers. This provides a novel approach, combining exploratory multivariate analysis (PCA and LDA) for biomarker extraction, with diagnostic evaluation through classifier simulation. Investigating primary changes in leaf tissue due to infection and disease progression, as well as classification of infection category based on class labels (time after infection) was intended to show the practical applicability of portable ATR-FTIR spectroscopy for the study of plant-pathogen interactions.

**Materials and Methods**

***Plant Growth Conditions***

Tomato plants, *Solanum lycopersicum* cv. Moneymaker (Thompson and Morgan, Ipswich, UK) were grown from seed in 1 L pots containing Levington’s M3 growth medium (Levington Horticulture Ltd, Ipswich, UK) for 3 weeks under glasshouse conditions (40-60% relative humidity, 23.5-28.5°C) with an 18/6 h day/night cycle (minimum illumination 500 µmol m-2 s-1 at the plant canopy from 600 W metal-halide lamps). Plants were watered daily between 8-9 am to holding capacity.

***Botrytis cinerea Inoculum Preparation and Infection of Plants***

Plants were inoculated with a suspension of spores of *B. cinerea* according to Asselbergh et al. (2007) and optimized for plants grown under the specified glasshouse conditions. Frozen (-80°C) 8 mL stock plugs of potato dextrose agar (PDA) containingmycelium of *B. cinerea* [strain R16] (Faretra and Pollastro 1991), were placed in the center of petri dishes containing freshly prepared PDA medium. The Petri dishes were incubated in a dark growth chamber (Percival AR-36L3) at 22°C and 100% humidity for 5 days until complete mycelial coverage of the PDA medium, after which they were exposed to a near-UV (UVA; 350-500 nm) light cycle (12 h dark/light; intensity: 28 µmol m-2 s-1; bulb: Fluora L 18 W/77, Osram, Munich, Germany) for 7 days to induce sporulation (Schumacher 2017). Loose spores were washed into a 50 mL falcon tube, using approx. 15 mL deionized water. The remaining PDA plate containing mycelium and spores was cut into pieces using a sterile scalpel and added to a separate falcon tube containing 0.01% Tween-20 (**Polyoxyethylenesorbitan monolaurate)** in 20 mL deionized water, and subsequently vortexed for 3 min. After intense mixing, the solution was gravity filtered through double-layered 20 µm nylon mesh into a 50 mL falcon tube to separate spores from mycelial debris and PDA medium. Filtrate containing spores in 0.01% aqueous Tween-20, as well as the separate loose spores, were centrifuged for 15 min at 15,000 relative centrifugal force (rcf) at 15°C. Supernatant was removed by decanting and spores were re-suspended and combined in 15 mL molecular grade water (Sigma Aldrich, St. Louis, Missouri, USA). The concentration of spores determined via hemocytometer was adjusted to 5×105 spores mL-1 in 0.05 M KH2PO4 and 33 mM glucose respectively (Asselbergh et al. 2007). This solution was prepared 3 h before application to allow pre-germination of spores, prior to infection of plants. Individual tomato plants were briefly removed from the glasshouse, placed into a containment area where only the shoot was exposed, and uniformly sprayed from above with approx. 1 ml of spore solution at approx. 45° from plant canopy at 20 cm distance. This was repeated four times rotating plants 45° after each. Control plants were treated with a mock solution containing only 0.05 M glucose, 33.3 mM KH2PO4, and no spores. Following inoculation, plans were returned to the glasshouse to promote infection. Humidity was maintained at 100% for 24 h using a glasshouse mister (Easy Irrigation, UK) combined with water timer (Easy-Control 1882, Gardena, Ulm, Germany) producing a 50 µm droplet diameter spray for 15 min every 2 h. Twenty plants were used for each treatment (20 mock/20 infected); 19 out of 20 plants infected developed symptoms, while the 20 mock controls remained asymptomatic throughout the experiment. Eight plants from each treatment (mock and infected) were reserved to confirm symptom development independent of ATR-FTIR analysis. Symptoms were assessed visually at the three measured time points and used as class criteria. Detailed description of symptoms is summarized in SI Table 1 and shown in Figure 1. Described symptoms and time of onset after exposure to fungal pathogen were consistent with various stages of infection as previously described (Asselbergh et al. 2007; Audenaert et al. 2002; El Oirdi et al. 2011) and thus covered the range of disease progression relevant for crop protection.

***ATR-FTIR Spectroscopy***

Vibrational spectra from tomato plant leaves were taken using a Bruker Alpha IR spectrometer with Platinum ATR attachment (Bruker Optics, UK). Spectra were acquired over the range 4000-400 cm-1 with a spectral resolution of 8 cm-1, 32 co-additions and a mirror velocity of 7.5 kHz for optimum signal to noise ratio and acquisition speed. Background spectra were taken prior to each sample to account for the ambient atmosphere. The diamond ATR crystal defined a spatial resolution (sampling area) of 1 mm2. ATR cleaning wipes containing isopropyl-alcohol (Bruker, UK) were used to clean the ATR diamond crystal between sample measurements.

A total of 16 leaves (4 controls, 4 infected plants) were measured for the three different categories (pre-symptomatic, intermediate symptomatic, advanced symptomatic). Ten spectra were collected from each leaf, resulting in 80 spectra for each category (240 total spectra ). Five spectra were taken from each of the two main leaflets, comprising each biological replicate (plant). Four plants (n=4) were measured for controls and infected plants respectively at each timepoint. Spectra taken from symptomatic leaves at 96 and 144 h post infection were from regions of remaining healthy tissue (for an example of acquisition points on healthy and infected leaves, see Figure 1). Plants were moved briefly to the laboratory but maintained at 25.0-27.5°C and 50% relative humidity under a portable 600 W high-pressure sodium lamp (Omega lighting, Berkeley, CA) source (minimum of 500 µmol m-2 s-1) during spectral acquisition to simulate near to glasshouse conditions during spectroscopic analysis.

***Computational Analysis***

All pre-processing and data analysis were carried out using the PLS toolbox version 7.9 (Eigenvector Research, Inc., WA, USA) in conjunction with MATLAB 2016a (The Math Works, MA, USA). Pre-processing of raw spectra was performed by selecting the MIR fingerprint region (1800–900 cm-1), baseline correction using automatic weighted least squares (AWLS) followed by vector normalization. Initially, the mean-centered pre-processed spectral data were evaluated by means of principal component analysis (PCA) (Bro and Smilde 2014). PCA is an exploratory analysis method that reduces the original variables (*i.e.*, wavenumbers) to a few number of principal components (PCs) accounting for the majority of the original data variance. Each PC is orthogonal to each other, where the first PC accounts to the maximum explained variance followed by the second PC and so on. The PCs are composed of scores and loadings, where the first represents the variance on sample direction, thus being used to assess similarities/dissimilarities among the samples; and the latter represents the contribution of each variable for the model decomposition, thus being used to find important spectral markers. PCA was the method of choice for analyzing this dataset since it is simple, fast, and combines exploratory analysis, data reduction, and feature extraction into one single method. PCA scores were used to explore overall dataset variance and any clustering related to plant infection independent of class label, while the loadings on the first two PCs were used to derive specific biomarkers indicative of plant infection at different stages (infection category). PCA was used to explore overall dataset variance, which was related to plant infection independent of class label. This was to determine whether the dataset variance was caused primarily by the infection (observed symptoms) and to rule out other underlying effects such as naturally occurring tissue heterogeneity.

Discriminant analysis was performed to distinguish the samples into control and infected in a predictive multivariate fashion. This was achieved by means of principal component analysis linear discriminant analysis (PCA-LDA). In PCA-LDA, a linear discriminant analysis (LDA) classifier is employed in the PCA scores (Morais et al. 2018). LDA finds the best linear discriminant direction between the groups, maximizing the distance between the classes and minimizing the samples distance within each class. The PCA-LDA classification score for each sample ($cf(t\_{i})$) can be estimated in a non-Bayesian form by a Mahalanobis distance calculation as follows (Morais et al. 2018):

$cf(t\_{i})=\left(t\_{i}-\overbar{t}\_{k}\right)^{T}C\_{pooled}^{-1}\left(t\_{i}-\overbar{t}\_{k}\right)$ (1)

where $t\_{i}$ is a vector containing the PCA scores for all selected PCs for a given sample *i*; $\overbar{t}\_{k}$ is the mean PCA scores vector for class *k*; $C\_{pooled}$ is the pooled covariance matrix; and the superscript T represents the matrix transpose operation. All spectra were used for PCA-LDA. To prevent having to average spectra, five spectra were taken to measure different locations on each leaflet to account for intra-variability across the measured leaf surface. Spectra from different locations on the leaflets provided point-based but robust model input. PCA-LDA models were built using 70% of the samples in the training set, where the number of selected PCs was optimized *via* 10-k fold cross-validation. This cross-validation method uses 10% of the data for a temporary validation set. These spectra do not belong to the training set or the initial PCA model, and their PCA scores are calculated by using the PCA loadings of the training samples by an inverse-matrix calculation and are used to assess real model performance. All results expressed in this manuscript use cross-validation in order to eliminate the risk of overfitting (Siqueira et al. 2018). Overfitting is observed by a difference in the model performance by comparing the training and validation (or cross-validation) sets where, in presence of overfitting, the latter tends to generate much higher errors (Hibbert 2016; Siqueira et al. 2018). The model validation was performed with 30% of the samples, where metrics such as accuracy, sensitivity and specificity were estimated. The sample splitting into training and validation sets was performed using the Kennard-Stone algorithm (Kennard and Stone 1969). For this dataset, four PCs were sufficient to account for more than 95% cumulative variance in each case (SI Table 2).

**Results and Discussion**

***Botrytis cinerea Infection Induces Spectral Alterations in Tomato Plant Leaves***

Spectral alterations are observed in leaf tissue of *B. cinerea* infected tomato plants compared to healthy control plants. Primary absorbance peaks of tomato leaves were determined based on the class means from pairwise comparisons of control and infected tissue over time (Figure 2). These show the strongest peaks in the carbohydrate fingerprint region (1200-900 cm-1), as well as the upper fingerprint (UF) region from 1800-1500 cm-1.While the UF region is generally associated with strong protein vibrations, this region also contains vibrations from water, polysaccharides, lipids, nucleic acids, as well as various other compounds (Berthomieu and Hienerwadel 2009; Movasaghi et al. 2008). Water absorbs strongly in the region around 1600 cm-1, which is part of the fingerprint region under investigation. It becomes apparent however from the class mean spectra comparisons between control and infected plants that the broad peak including the region 1600 cm-1 is variable and not decreasing in infected plants compared to controls. This is inconsistent with observed water loss caused by cuticle breakdown and tissue degradation due to infection. It is therefore likely that this region is rather a mixture of plant compounds that primarily include biomolecules, and where water appears to have a minimal influence on the spectrum. The study by Butler et al. (2015) also employed the use of ATR-FTIR to circumvent the effects of water on IR measurements, and analysed plant leaves of tomato.

In plants, the region between 1800-1500 cm-1 potentially reflects various compounds including the primary group of biomolecules carbohydrates, proteins, lipids, and nucleic acids (Ord et al. 2016; Ribeiro da Luz 2006). Concentrations of these biomolecules can vary substantially in leaves and throughout the plants life cycle but are in the lower microgram per gram of fresh leaf weight for nucleic acids; in the range of up to milligrams per gram for lipids and proteins (Novitskaya et al. 2000;Smillie and Krotkov 1961). Carbohydrates in the form of starch, free sugars, and cellulose contribute to the spectral signature of plant leaves and may be found at higher concentrations. Starch and soluble sugars are generally in the milligram per gram range, while cellulose may be in the hundreds of milligrams per gram due to its structural prominence (Li et al. 2015; Wilson 1965). Yet the spectrum of intact leaf is a complex combination of biomolecules, most of which are expected to come from surface structures such as the large proportion of cellulose and cuticle components as previously characterized (Dominguez et al. 2011; Heredia-Guerrero et al. 2014; Largo-Gosens et al. 2014). *In vivo*,epidermal surface compounds primarily contribute to the tomato leaf spectrum, which is strongly influenced by structural polymers such as cellulose, lignin, and Cutin vibrations abundant in the cutinized cell wall of plant epidermis (Heredia-Guerrero et al. 2014; Largo-Gosens et al. 2014; Skolik et al. 2018).

While overall absorbance is lower over the region 1500-1200 cm-1, several distinct peaks are nonetheless evident. Figure 2a-c shows the unprocessed class mean spectra for healthy control plants, versus plants exposed to *B. cinerea*; pre-processed class means (baseline corrected and normalized) are shown in Figure 2d-f for comparison (see SI Figure 1 original raw spectra). Comparison of healthy and infected tomato leaves shows spectral variations in the pre-symptomatic (PS) stage at 48 h (Figure 2a and d), at which point plants do not yet show any visual symptoms associated with *B. cinerea* infection (Figure 1, 48 h PS). Clear differences at PS stage are only observed in the region from 1800-1500 cm-1 (Figure 2d), but not over the carbohydrate region between 1200-900 cm-1, which become clear at the intermediate symptomatic (IS) at 96 h and advanced symptomatic (AS) at 144 h (Figure 2e and f). Importantly, plants measured at the PS stage developed visual symptoms at later stages, consistent with those observed at IS and AS, confirming the presence of *B. cinerea* at the PS stage. From class mean spectra, clear alterations in the region between 1500-1200 cm-1 are not observed until 144 h (Figure 2f). Specifically, differences in the carbohydrate fingerprint region are not observed PS, although slight variation is seen between 1750-1400 cm-1 (Figure 2d). Spectral divergence like that seen PS between 1750-1400 cm-1, is shifted to between 1650-1300 cm-1 IS and clear alterations to the carbohydrate region between 1200-900 cm-1 emerge at the IS stage (compare Figure 2d and e). The differences in these regions, indicate important alterations in the underlying biochemical architecture of leaf tissue caused by pathogen infection, which is evident from direct comparison of mean fingerprint spectra of control and infected tomato plants. Increased spectral shifts clearly coincide with the development of visual symptoms on leaves. As infection with *B. cinerea* progresses, characteristic tissue changes occur including the described development of large lesions covering most of the leaf area, black necrotic tissue, visible hyphae and conidiophores indicating pathogen sporulation (Figure 1 and SI Table 1).

Increasing changes in the carbohydrate and UF regions of plants affected by *B. cinerea* suggest structural changes in plant leaf polysaccharides, proteins, lipids and waxes, as well as potentially nucleic acids. A clear decrease in the absorption over the carbohydrate fingerprint region (1200-900 cm-1) during the infection process, is associated with changes in biochemical composition of polysaccharides such as pectin and cellulose, as well as cutin, which are present in the surface layers of the epidermis and expected to arise mainly from changes to the cuticle and cell wall (Largo-Gosens et al. 2014; Heredia-Guerrero et al. 2014). Changes in carbohydrates, proteins, and moisture content, are consistent with processes such as cuticle and cell wall degradation resulting in water loss, following pathogen invasion (Asselbergh et al. 2007; Audenaert et al. 2002; El Oirdi et al. 2011). The two main regions, 1750-1500 and 1200-900 cm-1 clearly show the largest differences and may be of considerable importance in the detection of leaf tissue modifications associated with necrotrophic pathogen attack *in vivo* based on MIR spectra, where the UF and carbohydrate sections of the MIR spectrum show the largest variation (Figure 2). The two regions combined, collectively contain specific compounds that change in response to pathogen invasion. Specifically, the UF region contains compounds such as polysaccharides including pectin but to a lesser extent than the carbohydrate region. It is additionally associated with lignin or related compounds and secondary metabolites all present in plant surface structures (Largo-Gosens et al. 2014; Monti et al. 2013; Schulz and Baranska 2007). Changes to the bulk of these compounds also occur during infection with pathogenic fungi (Asselbergh et al. 2007; Heredia-Guerrero et al. 2014; Largo-Gosens et al. 2014). Although the region between 1500-1200 cm-1 showed only slight changes in the class mean spectra over the infection period, it may present an alternative target region to probe the subtler changes occurring *B. cinerea* infection, which can only be reliably determined through multivariate analysis of the spectral data.

***Disease Progression Generates Unique Spectral Profiles and Specific Biomarkers at Distinct Stages of Botrytis cinerea Infection***

*B. cinerea* infection generates spectral profiles unique to PS, IS, and AS stages of disease progression. Based on the observable differences in the class mean spectra, unsupervised PCA analysis was performed which showed that *B. cinerea* infection was responsible for at least 85% of the variance observed in the spectral data, using the described data analysis approach. For this dataset, four PCs were enough to account for a minimum of 95% total variance for each group (SI Table 2).

Consistent separation of spectral classes (infection categories PS, IS, and AS) was observed along the first two PCs, which accounted for a minimum of 84.89%% of overall dataset variance. At the PS stage, separation along PC1 accounted for 65.52% of dataset variance, while PC2 accounted for 27.96% accounting for a total of 93.48% variance and producing the best class separation using PCA (Figure 3a). Spectra from IS and AS plants also showed clear separation from their respective controls along PC1 and PC2 (Figure 3b and c). For plants showing IS, class separation along PC1 and PC2 explained 73.97% and 10.92% of dataset variance, respectively, while AS plants showed PC variance at 63.18% (PC1) and 22.56% (PC2), explaining a total variance of 84.89% and 85.72% respectively. Better class separation and accountancy of variance suggests that biochemical changes were most pronounced in PS stages of plant infection, based on results of unsupervised component analysis. For PS detection of plant disease, it is encouraging that the best separation is achieved for the PS category where no visual symptoms are apparent in contrast to leaves clearly infected and colonized by *B. cinerea* at later stages (IS and AS).

To associate vibrational modes with dataset variance and the observed separation of infected classes based on PC1 and PC2 scores (Figure 3a-c), PCA loadings for PC1 and PC2 were generated (Figure 3d-f). Major wavenumber peaks within PC loadings represent the main vibrational modes, and thus functional groups and associated compounds, altered because of infection. Compared to PCA scores (Figure 3a-c), loadings provide highly specific ‘wavenumber biomarkers’ responsible for spectral variations between healthy and diseased plants (Kelly et al. 2011). Table 2 summarizes these tentatively assigned spectral biomarkers for PS, IS, and AS (Figure 3d, e, and f, respectively). Similar wavenumber biomarkers identified over various regions of the fingerprint spectrum are common to all disease stages, where only a small fraction of wavenumbers, specifically 1748, 1645, 1580, 975, and 945 cm-1 are uniquely assigned to a specific category (Table 1; **bold** wavenumbers). Aside from these unique biomarkers similar compounds are identified at all stages of disease progression. Compounds generally related to *B. cinerea* infection are identified over the regions 1600-1590, 1539-1537, 1149-1148, 1105-1088, 1063-1061, 1038-1037, and 1000-995 cm-1, which were consistent with alterations in the two strongest absorbing UF (1800-1500 cm-1) and carbohydrate (1200-900 cm-1) regions identified from the class mean spectra in Figure 2. The large degree of overlapping spectral markers between PS, IS, and AS plants indicates alterations in similar biochemical compounds through disease progression over time and thereby links biomarkers from each distinct disease stage.

Compounds linking disease stages, extracted through multivariate analysis, belong predominantly to carbohydrates, proteins, nucleic acids, plant polymers, and secondary metabolites (Table 1). Main changes over the UF region included mainly protein vibrations including protein phosphorylation, vibrations associated with pectin and other polysaccharides, lignin, as well as an abundance of compounds associated with secondary plant metabolites. Table 1 also shows that most of the compounds in the UF region relate to structural polymers including various vibrations associated with pectin. Pectin decomposition is thought to be necessary for *Botrytis* proliferation on all hosts (Asselbergh et al. 2007) and may therefore become one of many targets for spectroscopic investigations. Quantitative analysis of spectral biomarkers would permit the investigation of positive or negative correlations between individual compounds such as pectin for example. At the other end of the spectrum, the carbohydrate region shows alterations predominantly in polymers such as cellulose, pectin, polysaccharides, including some specific polysaccharides such as arabinogalactan and xyloglucan related entities (Largo-Gosens et al. 2014). Such entities along with the carbohydrates mentioned may be part of carbohydrate partitioning, which via invertase enzymes may be linked to *B. cinerea* infection in tomatoes (Hyun et al. 2011). Further, this mechanism may be linked to plant defense through the jasmonic acid signaling pathway (Hyun et al. 2011). Connections like these may ultimately link spectral biomarkers with plant physiology / pathology reliably. While the extensive evaluation of all generated spectral biomarkers requires additional studies, it appears that there is a potential for links between alterations in spectral biomarkers and plant disease.

Vibrational modes at 1645, 1539, and 1038 cm-1 may contain absorbance peaks generated directly by the fungal pathogen, although these have not been assigned to nucleic acids previously (Salman et al., 2010, 2012). Contribution to absorbance at these wavenumber regions may come from fungal constituents including proteins (1645 and 1539 cm-1) and chitin at 1038 cm-1 (Salman et al., 2010, 2012; Skolik et al. 2019). In general, it remains to be determined to which degree the *B. cinerea* secretome and its physical presence, growth, and spread influences spectral analysis and pathogen detection. In so far, the ATR-FTIR spectral data represents the complex interaction of tomato leaf tissue and fungal mass as a mixture effect and further research will determine in detail the contributions of each organism to the observed spectral changes.

A small fraction of wavenumbers 1748, 1645, 1580, 975, and 945 cm-1 were unique to PS, IS, and AS. Categorically, 1645, 1580, and 975 cm-1 were unique markers to the PS disease stage, while 1748 and 945 cm-1 were associated with IS and AS, respectively (Table 2). In contrast to IS and AS stages, the PS stage showed the highest number of unique markers, compared to only one for both IS and AS stages. This is consistent with the higher degree of variance explained for PS along PC1 and PC2 compared to IS and AS plants. While PS markers covered both upper (1645 and 1580 cm-1) and lower (975 cm-1) regions of the fingerprint spectrum, IS showed unique absorbance only at 1748 cm-1, whereas AS was uniquely identified by absorbance at 945 cm-1. Unique fingerprint changes in the PS stage were associated with changes in the UF and carbohydrate regions, specifically the Amide I region of proteins (1645 and 1580 cm-1) and polysaccharides associated with pectin (975 cm-1) (Abidi et al. 2014; Butler et al. 2015; Butler et al. 2017; Monti et al. 2013; Ord et al. 2016; Schulz and Baranska 2007). Alternatively, these regions can also be related to changes in phenolic compounds of secondary metabolites, and protein phosphorylation (Heredia-Guerrero et al. 2014; Ord et al. 2016; Schulz and Baranska 2007). Both alterations to pectin content and structure, as well as changes to proteins, including protein phosphorylation occur during *B. cinerea* infection in both plant and pathogen (Asselbergh et al. 2007; Ehness et al. 1997; El Oirdi et al. 2011; Segmüller et al. 2007). Additionally, secondary metabolites including terpenes were identified, which may contribute to the MIR spectrum; these are known to be employed to deter pathogens (Gershenzon and Dudareva 2007; Zengin and Baysal 2014).

The IS stage showed a single unique wavenumber at 1748 cm-1. This wavenumber corresponds predominantly to changes in carbohydrates and or lipids, including pectin and waxes, likely reflecting substantial changes in plant epidermal layers such as the cuticle and cell wall (Abidi et al. 2014; Butler et al. 2017; Largo-Gosens et al. 2014). At this stage, changes in pectin appear to continue, consistent with pectin degradation during *B. cinerea* infection, which may be influenced through the abscisic acid pathway that influences pectin stability and thereby susceptibility to the pathogen (Asselbergh et al. 2007; El Oirdi et al. 2011). Changes in waxes and related compounds could be indicative of damage to the cuticle, which is visually evident at the IS stage (Figure 1). Onset of visual symptoms, at both IS and AS, are likely linked to processes including programmed cell death (PCD) and tissue necrosis (Dieryckx et al. 2015; Hoeberichts et al. 2003).

At the other extreme of the fingerprint spectrum, a vibrational mode at 945 cm-1 is a unique spectral biomarker to the AS stage. Compounds related to this wavenumber include pectin, protein phosphorylation, and secondary metabolites, plausibly reflecting continued pectin degradation, alterations in protein structure through covalent modifications (phosphorylation) of either plant of pathogen proteins, and or changes in secondary metabolite production, as a response to the necrotrophic mode of *B. cinerea* (Largo-Gosens et al. 2014; Ord et al. 2016; Schulz and Baranska 2007). This biomarker may be closely related to the marker identified at 975 cm-1 in the PS stage, where covalent changes to proteins appear to be involved (Table 1).

Determining the exact identity of IR compounds and their related processes requires additional research as well as replicate experiments. Currently available methods do not permit reliable biochemical validation of entities exactly, which remains a significant drawback of spectroscopy-based approaches. For this reason, identifying target compounds relevant for plant-pathogen interactions with necrotrophic fungi is a vital first step in pathogen detection and generating disease specificity based on spectral biomarkers, especially within intact plant systems. The large number of compounds identified by MIR spectroscopy and multivariate analysis (Table 1) provides insight into several potential molecular mechanisms associated with *B. cinerea* infection. While the identity of these mechanisms remains unclear, compounds identified by ATR-FTIR spectroscopy are compounds whose changes are consistent with plant-pathogen interactions (Ehness et al. 1997; Gershenzon and Dudareva 2007; Hyun et al. 2011; Segmüller et al. 2007; Zengin and Baysal 2014).

***Detecting Disease Progression Directly by Discrimination and Classification of Infection Category***

To determine the effectiveness of direct pathogen detection based on plant tissue modifications, classification algorithms may be especially useful as they do not rely on the validation of biochemical entities to be effective. This dual approach, combining exploratory with diagnostic data analysis, provides data that may help answer specific biological questions as well as immediate discriminatory information for detecting plant disease autonomously (Kelly et al. 2011; Trevisan et al. 2012).

Direct detection of plant infection throughout disease progression can be achieved autonomously by combining spectral data with multivariate analysis and machine classifiers. Biochemical information in the form of MIR spectra has been combined previously with supervised multivariate analysis and diverse classification algorithms to assess computer-based decision-making for autonomous diagnostics performance using MIR spectral input for disease detection in animal models (Martin et al. 2007; Morais and Lima 2017; Trevisan et al. 2012). Complete class separation was attained for PS, IS, and AS plants after exposure to *B. cinerea*; Figure 3 shows PCA-LDA scores plots and clear separation of healthy plants from PS (A), IS (B), and AS (C) across DF1. Based on the complete data separation observed along DF1, diagnostic potential for plant disease detection was evaluated via classifier output (Table 2). PCA-LDA classification accuracy was equal to 100% in the validation set and was thus highly effective for disease detection for all groups (PS, IS, AS). All groups of plants were classified with 100% sensitivity and specificity. The models were trained with 4 PCs (SI Table 2), having a training accuracy of 100% for all types of infection comparisons (Table 2), and a cross-validation accuracy of 98% for IS and 100% for both PS and AS (SI Table 3).

Near perfect classification rates strongly suggest that biospectroscopy, combined with chemometric classifiers, offers a suitable approach to the development of diagnostic tools for plant disease. Importantly, the accurate classification of PS plants demonstrates that early identification of *B. cinerea* infection can be achieved in completely intact tomato plants, prior to the onset of symptoms on which current diagnostics approaches rely. Therefore, detection of PS plants is especially valuable, because the PS disease stage offers a longer window for the application of crop protection measures, which are becoming part of specialized modern approaches including integrated pest management (IPM) and precision crop protection (Barzman et al. 2015).

***Conclusions and Perspectives***

MIR biospectroscopy is a capable technology applicable to the development of precision sensors aimed at mitigating crop loss due to pests and pathogens. However, although biospectroscopy presents a novel analytical technique, it has as yet been little used in the plant and crop sciences, especially for applications to dynamic intact plant-pathogen systems (Skolik et al. 2019). Here it was demonstrated that biospectroscopy in the form of semi-portable ATR-FTIR was effective at non-destructive *in vivo* analysis of plant-pathogen interaction between *B. cinerea* and *S. lycopersicum* at the leaf level (Figure 1). Clear differences in the class mean spectra were observed primarily in the UF and carbohydrate regions (Figure 2). This was consistent with specific spectral biomarkers extracted via multivariate analysis by way of cluster analysis and through PCA loadings (Figure 3 and Table 1). This revealed that the majority of variance within the spectral data of the MIR fingerprint were due to plant infection. The main biochemical variations and their potential compound identity were tentatively assigned and discussed. Main candidate targets identified included carbohydrates, specifically pectin, proteins and protein phosphorylation, lipids and waxes, as well as secondary metabolites. Additional targets include nucleic acid and more subtle compounds in the region between 1500-1200 cm-1, which have yet to be explored in depth. In summary, most predominant modifications were detected in the spectral UF and carbohydrate regions, which were consistent with changes occurring in plant leaves because of colonization and attack by *B. cinerea* including later stage tissue degradation and necrosis (Asselbergh et al. 2007; El Oirdi et al. 2011). A next step to link spectral data with plant physiology is the quantitative assessment of specific compounds, such as pectin, in relation to disease progression, in order to establish a firm link between MIR spectroscopy and biological processes. Supervised PCA-LDA analysis completely segregated infected from non-infected plants at PS, IS, and AS stages of plant disease, showing promisingly high classification accuracy for applied disease detection (Figure 4 and Table 2). With a classification accuracy of 100%, this approach appears highly suitable for pre- and post-symptomatic disease detection.

Rapid sensor-based disease detection will contribute to reductions in crop loss and increase food security overall, by facilitating the optimization of crop protection products, limit their overuse, while also reducing human and environmental exposure to harmful chemicals. However, further research is required with respect to MIR biospectroscopy-based disease detection to increase the range of studies performed on intact plants *in vivo* focusing on model plants/crops, as well as the evaluation of portable equipment potentially suitable for the field (Skolik et al. 2018). Additionally, slight re-tooling of currently available MIR spectroscopy equipment, will permit further proof-of-concept trials to be instigated in the near future as has recently been achieved through the use of portable and handheld Raman spectrometers (Egging et al. 2018; Farber and Kurouski 2018; Yeturu et al. 2016). An unexplored aspect of MIR biospectroscopy is the use of acquisition modes for liquid and gaseous samples, which to date remain virtually unexplored, but offer additional potential for disease detection and plant-environment interactions relevant to crop biology (Skolik et al. 2018). While the spectrochemical analysis of intact plant-pathogen systems is still in the beginning stages, the rapid growth of this field and the largely untapped potential of this technology will ensure its future contribution to the fields of plant and crop science.

**References**

Abidi, N., L. Cabrales, and C.H. Haigler. 2014. Changes in the cell wall and cellulose content of developing cotton fibers investigated by FTIR spectroscopy. *Carbohydrate Polymers* 100:9-16. doi: 10.1016/j.carbpol.2013.01.074

Aschemann-Witzel, J. 2016. Waste not, want not, emit less. *Science* 6284: 408-409. doi: 10.1126/science.aaf2978

Asselbergh, B., K. Curvers, S-C. França, K. Audenaert, M. Vuylsteke, F. Van Breusegem, and M. Höfte. 2007. Resistance to Botrytis cinerea in sitiens, an abscisic acid-deficient tomato mutant, involves timely production of hydrogen peroxide and cell wall modifications in the epidermis. *Plant Physiology* 144: 1863-1877. doi: 10.1104/pp.107.099226

Audenaert, K., G-B. De Meyer, and M. Höfte. 2002. Abscisic acid determines basal susceptibility of tomato to Botrytis cinerea and suppresses salicylic acid-dependent signaling mechanisms. *Plant physiology* 128: 491-501. doi: 10.1104/pp.010605

Barzman, M., P. Bàrberi, A.N.E. Birch, P. Boonekamp, S. Dachbrodt-Saaydeh, B. Graf, B. Hommel et al. 2015. Eight principles of integrated pest management. *Agronomy for sustainable development* 35: 1199-1215. doi: 10.1007/s13593-015-0327-9.

Behmann, J., A.-K. Mahlein, T. Rumpf, C. Römer, and L. Plümer. 2015. A review of advanced machine learning methods for the detection of biotic stress in precision crop protection. *Precision Agriculture* 16: 239-260. doi: 10.1007/s11119-014-9372-7

Berthomieu, C., and R. Hienerwadel. 2009. Fourier transform infrared (FTIR) spectroscopy. *Photosynthesis research* 101: 157-170. doi: 10.1007/s11120-009-9439-x

Bourguet, D., and T. Guillemaud. 2016. The hidden and external costs of pesticide use. In *Sustainable agriculture reviews*, ed. E. Lichtfouse , vol 19, 1st ed., 35-120, Springer.

Butler, H.-J., Steven A., Martin R.-M., and F.-L. Martin. 2017. Detecting nutrient deficiency in plant systems using synchrotron Fourier-transform infrared microspectroscopy. *Vibrational Spectroscopy* 90:46-55. doi: 10.1016/j.vibspec.2017.03.004

Butler, H.-J., M.-R. McAinsh, S. Adams, and F.-L. Martin. 2015. Application of vibrational spectroscopy techniques to non-destructively monitor plant health and development. *Analytical Methods* 7: 4059-4070. doi: 10.1039/c5ay00377f

Chan, KL-A., and S.-G. Kazarian. 2016. Attenuated total reflection Fourier-transform infrared (ATR-FTIR) imaging of tissues and live cells. *Chemical Society Reviews* 45: 1850-1864. doi: 10.1039/C5CS00515A

Dayan, F.-E., C.-L. Cantrell, and S.-O. Duke. 2009. Natural products in crop protection. *Bioorganic and Medicinal Chemistry* 17: 4022-4034. doi: 10.1016/j.bmc.2009.01.046

Dieryckx, C., V. Gaudin, J-W. Dupuy, M. Bonneu, V. Girard, and D. Job. 2015. Beyond plant defense: insights on the potential of salicylic and methylsalicylic acid to contain growth of the phytopathogen Botrytis cinerea. *Frontiers in Plant Science* 6: 859. doi: 10.3389/fpls.2015.00859

Domínguez, E., J-A. Heredia‐Guerrero, and A. Heredia. 2011. The biophysical design of plant cuticles: an overview. *New Phytologist* 189: 938-949. doi: 10.1111/j.1469-8137.2010.03553.x

Egging, V., J. Nguyen, and D. Kurouski. 2018. Detection and identification of fungal infections in intact wheat and sorghum grain using a hand-held Raman spectrometer. *Analytical Chemistry* 90: 8616-8621. doi: 10.1021/acs.analchem.8b01863

Ehness, R., M. Ecker, D-E. Godt, and T Roitsch. 1997. Glucose and stress independently regulate source and sink metabolism and defense mechanisms via signal transduction pathways involving protein phosphorylation. *The Plant Cell* 9: 1825-1841. doi: 10.1105/tpc.9.10.1825

Elad, Y., and A. Stewart. 2007. Microbial control of Botrytis spp. In *Botrytis: biology, pathology and control*, ed. Y. Elad, B. Williamson, P. Tudzynski, and N. Delen, vol. 1, 1st ed., 223-241. Springer.

El Oirdi, M., T.-A. El Rahman, L. Rigano, A. El Hadrami, M.-C. Rodriguez, F. Daayf, A. Vojnov, and K. Bouarab. 2011. Botrytis cinerea manipulates the antagonistic effects between immune pathways to promote disease development in tomato. *The Plant Cell* 23: 2405-2421. doi :10.1105/tpc.111.083394

Farber, C., and D. Kurouski. 2018. Detection and identification of plant pathogens on maize kernels with a hand-held Raman spectrometer. *Analytical Chemistry* 90: 3009-3012. doi: 10.1021/acs.analchem.8b00222

Faretra, F., and S. Pollastro. 1991. Genetic basis of resistance to benzimidazole and dicarboximide fungicides in Botryotinia fuckeliana (Botrytis cinerea). *Mycological Research* 95: 943-951. doi: 10.1016/S0953-7562(09)80091-9

Gershenzon, J., and N. Dudareva. 2007 The function of terpene natural products in the natural world. *Nature chemical biology* 3: 408-414. doi: 10.1038/nchembio.2007.5

Heredia-Guerrero, J.-A., J.-J. Benítez, E. Domínguez, I.-S. Bayer, R. Cingolani, A. Athanassiou, and A. Heredia. 2014. Infrared and Raman spectroscopic features of plant cuticles: a review. *Frontiers in Plant Science* 5: 305. doi: 10.3389/fpls.2014.00305

Hibbert, D.-B. 2016. Vocabulary of concepts and terms in chemometrics (IUPAC Recommendations 2016). *Pure and Applied Chemistry* 88: 407-443. doi: 10.1515/pac-2015-0605

Hoeberichts, F.-A., A.-T. Have, and E.-J. Woltering. 2003. A tomato metacaspase gene is upregulated during programmed cell death in Botrytis cinerea-infected leaves. *Planta* 217: 517-522. doi: 10.1007/s00425-003-1049-9

Hyun, T.-K., S.-H. Eom, Y. Rim, and J-S. Kim. 2011. Alteration of the expression and activation of tomato invertases during Botrytis cinerea infection. *Plant Omics* 4: 413-417.

Kelly, J.-G., J. Trevisan, A.-D. Scott, P.-L. Carmichael, H.-M. Pollock, P.-L. Martin-Hirsch, and F.-L. Martin. 2011. Biospectroscopy to metabolically profile biomolecular structure: a multistage approach linking computational analysis with biomarkers. *Journal of Proteome Research* 10: 1437-1448. doi: 10.1021/pr101067u

Kennard, R.-W., and L.-A. Stone. 1969. Computer aided design of experiments. *Technometrics* 11, no. 1: 137-148. doi: 10.1080/00401706.1969.10490666

Kimura, S., and N. Sinha. 2008. Tomato (Solanum lycopersicum): a model fruit-bearing crop. *Cold Spring Harbor Protocols* 11: pdb-emo105. doi: 10.1101/pdb.emo105

Largo-Gosens, A., M. Hernández-Altamirano, L. García-Calvo, A. Alonso-Simón, J. Álvarez, and J.-L. Acebes. 2014. Fourier transform mid infrared spectroscopy applications for monitoring the structural plasticity of plant cell walls. *Frontiers in plant science* 5: 303. doi: 10.3389/fpls.2014.00303

Lechenet, M., V. Bretagnolle, C. Bockstaller, F. Boissinot, M-S. Petit, S. Petit, and N.-M. Munier-Jolain. 2014. Reconciling pesticide reduction with economic and environmental sustainability in arable farming. *PloS one* 9: e97922. doi: 10.1371/journal.pone.0097922

Lechenet, M., F. Dessaint, G. Py, D. Makowski, and N. Munier-Jolain. Reducing pesticide use while preserving crop productivity and profitability on arable farms. *Nature Plants* 3: 1-6. doi: 10.1038/nplants.2017.8

Li, T., E. P. Heuvelink, and L.-FM Marcelis. 2015. Quantifying the source–sink balance and carbohydrate content in three tomato cultivars. *Frontiers in Plant Science* 6:416. doi: 10.3389/fpls.2015.00416

Martin, F.-L., M.-J. German, E. Wit, T. Fearn, N. Ragavan, and H.-M. Pollock. 2007. Identifying variables responsible for clustering in discriminant analysis of data from infrared microspectroscopy of a biological sample. *Journal of Computational biology* 14: 1176-1184. doi: 10.1089/cmb.2007.0057

Martin, F.-L., J.-G. Kelly, V. Llabjani, P.-L. Martin-Hirsch, I.-I. Patel, J. Trevisan, N.-J. Fullwood, and M.-J. Walsh. Distinguishing cell types or populations based on the computational analysis of their infrared spectra. *Nature Protocols* 5: 1748-1760. doi: 10.1038/nprot.2010.133

Mahlein, A-K. 2016. Plant disease detection by imaging sensors–parallels and specific demands for precision agriculture and plant phenotyping. *Plant disease* 100: 241-251. doi: 10.1094/PDIS-03-15-0340-FE

Monti, F., R. Dell’Anna, A. Sanson, M. Fasoli, M. Pezzotti, and S. Zenoni. 2013. A multivariate statistical analysis approach to highlight molecular processes in plant cell walls through ATR FT-IR microspectroscopy: The role of the α-expansin PhEXPA1 in Petunia hybrida. *Vibrational Spectroscopy* 65: 36-43. doi: 10.1016/j.vibspec.2012.11.010

Morais, C.-LM, and K.-MG Lima. 2017. Comparing unfolded and two-dimensional discriminant analysis and support vector machines for classification of EEM data. *Chemometrics and Intelligent Laboratory Systems* 170: 1-12. doi: 10.1016/j.chemolab.2017.09.001

Morais, C.-LM, R.-F. Shore, M.-G. Pereira, and F.-L. Martin. 2018. Assessing binary mixture effects from genotoxic and endocrine disrupting environmental contaminants using infrared spectroscopy. *ACS Omega* 3: 13399-13412. doi: 10.1021/acsomega.8b01916

Movasaghi, Z., S. Rehman, and I. ur Rehman. 2008. Fourier transform infrared (FTIR) spectroscopy of biological tissues. *Applied Spectroscopy Reviews* 43: 134-179. doi: 10.1080/05704920701829043

Novitskaya, G. V., T. A. Suvorova, and T. I. Trunova. 2000. Lipid composition of tomato leaves as related to plant cold tolerance. *Russian Journal of Plant Physiology* 47:728-733. doi: 10.1023/A:1026646826048

Oerke, E.-C. 2006. Crop losses to pests. *The Journal of Agricultural Science* 144: 31-43. doi: 10.1017/S0021859605005708

Ord, J., H.-J. Butler, M.-R. McAinsh, and F.-L. Martin. 2016. Spectrochemical analysis of sycamore (*Acer pseudoplatanus*) leaves for environmental health monitoring. *Analyst* 141: 2896-2903. doi: 10.1039/c6an00392c

Ribeiro da Luz, B. 2006. Attenuated total reflectance spectroscopy of plant leaves: a tool for ecological and botanical studies. *New Phytologist*, 172: 305-318. doi: 10.1111/j.1469-8137.2006.01823.x

Salman, A., E. Shufan, I. Lapidot, A. Pomerantz, M. Huleihel, L. Tsror, R. Moreh, and S. Mordechai. 2012. Identification of fungal phytopathogens using Fourier transform infrared-attenuated total reflection spectroscopy and advanced statistical methods. *Journal of biomedical optics* 17: 017002. doi: 10.1117/1.JBO.17.1.017002

Salman, A., L. Tsror, A. Pomerantz, R. Moreh, S. Mordechai, and M. Huleihel. 2010. FTIR spectroscopy for detection and identification of fungal phytopathogenes. *Spectroscopy* 24: 261-267. doi: 10.3233/SPE-2010-0448

Sankaran, S., A. Mishra, R. Ehsani, and C. Davis. 2010. A review of advanced techniques for detecting plant diseases. *Computers and Electronics in Agriculture* 72: 1-13. doi: 10.1016/j.compag.2010.02.007

Schulz, H., and M. Baranska. 2007. Identification and quantification of valuable plant substances by IR and Raman spectroscopy. *Vibrational Spectroscopy* 43: 13-25. doi: 10.1016/j.vibspec.2006.06.001

Schumacher, J. 2017. How light affects the life of Botrytis. *Fungal Genetics and Biology* 106: 26-41. doi: 10.1016/j.fgb.2017.06.002

Segmüller, N., U. Ellendorf, B. Tudzynski, and P. Tudzynski. 2007. BcSAK1, a stress-activated mitogen-activated protein kinase, is involved in vegetative differentiation and pathogenicity in Botrytis cinerea. *Eukaryotic cell* 6: 211-221. doi: 10.1128/EC.00153-06

Singbo, A.-G., A.-O. Lansink, and G. Emvalomatis. 2015. Estimating shadow prices and efficiency analysis of productive inputs and pesticide use of vegetable production. *European Journal of Operational Research* 245: 265-272. doi: 10.1016/j.ejor.2015.02.042

Sinha, A.-K., M.-G. Hofmann, U. Römer, W. Köckenberger, L. Elling, and T. Roitsch. 2002. Metabolizable and non-metabolizable sugars activate different signal transduction pathways in tomato. *Plant physiology* 128: 1480-1489. doi: 10.1104/pp.010771

Siqueira, L.-FS., C.-LM. Morais, R.-F. Araújo Júnior, A.-A. de Araújo, and K.-MG. Lima. 2018. SVM for FT‐MIR prostate cancer classification: An alternative to the traditional methods. *Journal of Chemometrics* 32: e3075. doi: 10.1002/cem.3075

Skolik, P., M.-R. McAinsh, and F.-L. Martin. 2019. ATR-FTIR spectroscopy non-destructively detects damage-induced sour rot infection in whole tomato fruit. *Planta* 249: 925-939. doi: 10.1007/s00425-018-3060-1

Skolik, P., M.-R. McAinsh, and F.-L. Martin. 2018. Biospectroscopy for plant and crop science. In *Comprehensive Analytical Chemistry*, ed. J. Lopes and C. Sousa vol. 80, 1st ed., 15-49. Elsevier.

Smillie, R.M. and G. Krotkov. 1961. Changes in the dry weight, protein, nucleic acid, and chlorophyll contents of growing pea leaves. *Canadian Journal of Botany*, 39:891-900.

Trebolazabala, J., M. Maguregui, H. Morillas, A. de Diego, and J.-M. Madariaga. 2013. Use of portable devices and confocal Raman spectrometers at different wavelength to obtain the spectral information of the main organic components in tomato (Solanum lycopersicum) fruits. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 105: 391-399. doi: 10.1016/j.saa.2012.12.047

Trebolazabala, J., M. Maguregui, H. Morillas, A. de Diego, and J.-M. Madariaga. 2017. Portable Raman spectroscopy for an in-situ monitoring the ripening of tomato (Solanum lycopersicum) fruits. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 180: 138-143. doi: 10.1016/j.saa.2017.03.024

Trevisan, J., P.-P. Angelov, P.-L. Carmichael, A.-D. Scott, and F.-L. Martin. 2012. Extracting biological information with computational analysis of Fourier-transform infrared (FTIR) biospectroscopy datasets: current practices to future perspectives. *Analyst* 137: 3202-3215. doi: 10.1039/c2an16300d

Vasileiadis, V.-P. 2017. Economic sustainability: Less pesticide rarely causes loss. *Nature Plants* 3: 1-2. doi: 10.1038/nplants.2017.16

Vleeshouwers, V.-GAA., and R.-P. Oliver. 2014. Effectors as tools in disease resistance breeding against biotrophic, hemibiotrophic, and necrotrophic plant pathogens. *Molecular plant-microbe interactions* 27: 196-206. doi: 10.1094/MPMI-10-13-0313-IA

Williamson, B., B. Tudzynski, P. Tudzynski, and J.-AL. Van Kan. 2007. Botrytis cinerea: the cause of grey mould disease. *Molecular plant pathology* 8: 561-580. doi: 10.1111/j.1364-3703.2007.00417.x

Wilson, D. 1965. Nutritive value and the genetic relationships of cellulose content and leaf tensile strength in Lolium. *The Journal of Agricultural Science*, 65:285-292. doi: 10.1017/S002185960004884X

Wunderlich, S.-M., and N.-M. Martinez. 2018. Conserving natural resources through food loss reduction: Production and consumption stages of the food supply chain. *International Soil and Water Conservation Research* 6: 331-339. doi: 10.1016/j.iswcr.2018.06.002

Yeturu, S., P.-V. Jentzsch, V. Ciobotă, R. Guerrero, P. Garrido, and L.-A. Ramos. 2016. Handheld Raman spectroscopy for the early detection of plant diseases: Abutilon mosaic virus infecting Abutilon sp. *Analytical Methods* 8: 3450-3457. doi: 10.1039/c6ay00381h

Zengin, H., and A.-H. Baysal. 2014. Antibacterial and antioxidant activity of essential oil terpenes against pathogenic and spoilage-forming bacteria and cell structure-activity relationships evaluated by SEM microscopy. *Molecules* 19: 17773-17798. doi: 10.3390/molecules191117773

**Figures**

Figure 1. Examples of spectral acquisition points on leaves of whole tomato plants. Measurements were taken from healthy looking tissue of infected leaves as shown (this was only necessary in symptomatic tissue (infected leaves at 96 and 144 h).

Figure 2, Pairwise comparisons of control (black) and *B. cinerea* infected (red) spectra taken from leaves of tomato plants over the fingerprint region. Mean spectral classes at 48, 96, and 144 h (A-C) and corresponding pre-processed mean spectra (D-F).

Figure 3. PCA 2-dimensional scores plots (A-C) and corresponding loadings (D-F) of pairwise control (black) versus *B. cinerea* infected (red) tomato leaf spectra at 48 (A and D), 96 (B and E), and 144 (C and F) h.

Figure 4. Cross-validated PCA-LDA scores plots of pairwise control (black) versus *B. cinerea* infected (red) tomato leaf spectra at 48 (A), 96 (B), and 144 (C) h.

**Tables**

Table 1. Biomarkers specific to infection categories PS, IS, and AS, identified by primary peaks of PCA loadings, representative of the main biochemical differences between infected and control plants.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Infection Category** | **Component** | **Wavenumber (cm-1)** | **Vibrational Mode** | **Biochemical Compound** | **References** |
| **Pre-symptomatic** **(48 h PI)** | PC-1 (65.6%) | 1590 a | Amide Iνas(COO) ν(C=N)ν(NH2)  | ProteinsPectin (non-esterified)LigninNucleic acids | [5], [6], [8] |
| 1098 a | ν(C-O-C) ester ν(C-O) ν(C-C) νs(PO42-) | CutinCellulose Pectin (as acetyl-ester)PolysaccharidesDNA/RNA | [1], [2], [4], [5], [6], [8], [9]  |
| 1061 a | ν(C-O) ν(C-C)δ(C–OH) δ(O-CH)  | Cellulose (particularly C3-O3H secondary alcohols) PolysaccharidesPectin Secondary metabolites(monoterpenes from leaf) | [1], [3], [5], [6], [9]  |
| 995 a | ω(CH2) ν(C-O) ring stretching ν(C-C) ring breathing  | CelluloseSecondary metabolites (monoterpenes from fruit) | [5], [7], [9] |
| PC-2 (27.9%) | **1645** | Amide I (incl. secondary structure in β-sheet motifs)δ(O-H)ν(C=C)  | ProteinsPectinAdsorbed waterSecondary metabolites (ethylene, monoterpenes from fruit, sesquiterpenes) | [1], [2], [3], [6], [8], [9]  |
| **1580** | Amide I ν(C–C) ring stretch | ProteinsPhenolic compounds | [7], [8] |
| 1092 c | ν(C-O) νs(PO42-) | Pectin (also as acetyl-ester)CelluloseDNA/RNA | [1], [5], [8] |
| **975** | ν(OCH3)ω(CH2)ω(RH-C-C-RH) | PectinProtein phosphorylationSecondary metabolites (monoterpenes from leaf) | [6], [8], [9] |
|  |  |  |  |  |  |
| **Intermediate Symptomatic****(96 h PI)** | PC-1 (74.0%) | 1592 a | Amide Iνas(COO) ν(C=N)ν(NH2)  | ProteinsPectin (non-esterified)LigninNucleic acids | [5], [6], [8] |
| 1149 d | ν(C-O-C)  | Pectin (ring and its glycosidic linkage)Non-cellulosic carbohydratesArabinogalactan | [1], [5], [6], [9] |
| 1063 a | ν(C-O) ν(C-C)δ(C-OH) δ(O-CH)  | Cellulose (particularly C3-O3H secondary alcohols) PolysaccharidesPectin Secondary metabolites (monoterpenes from leaf) | [1], [3], [5], [6], [9] |
| 1000 a | ν(C-O) ν(C-C)ν(O-C-H)  | Cellulose Pectin  | [1], [5], [9] |
| PC-2 (10.9%) | **1748** | ν(C=O)   | Pectin (also as ester)PolysaccharidesLipidsWax and suberin-like compoundsSecondary metabolites (monoterpenes from leaf) | [1], [3], [5], [6], [9] |
| 1592 a | Amide Iνas(COO) ν(C=N)ν(NH2)  | ProteinsPectin (non-esterified)LigninNucleic acids | [5], [6], [8] |
| 1539 d | Amide II (incl. secondary structure in β-sheet motifs)ν(C=N) δ(NH2) | ProteinsLigninNucleic acids | [1], [3], [5], [7], [8] |
| 1105 a | ν(C-O) ν(C-C)νs(C-O-C) ester νs(PO42-)  | Pectin (as acetyl-ester)CellulosePolysaccharidesCutinDNA/RNA | [1], [2], [4], [5], [6], [8], [9] |
| 1038 d | ν(C-O) ν(C-C)ν(CCO)ν(O-CH3) ν(CH2OH) | CellulosePectin (also as acetyl-ester)PolysaccharidesXyloglucanArabinogalactan Galactan | [3], [5], [8], [9] |
|  |  |  |  |  |  |
| **Advanced Symptomatic****(144 h PI)** | PC-1 (63.2) | 1600 a | ν(COO)ν(C=C)ν(C-C)ν(C=N)ν(NH2)  | PectinLignin Secondary metabolites (aromatic and phenolic compounds)Nucleic acids | [3], [4], [5], [6], [8], [9] |
| 1148 d | ν(C-O-C)  | Pectin (ring and its glycosidic linkage)Non-cellulosic carbohydratesArabinogalactan | [1], [5], [6], [9] |
| 1088 c | νs(C–O–C) | PectinXyloglucanArabinogalactan Secondary metabolites (monoterpenes from leaves) | [5], [9] |
| 1063 a | ν(C-O) ν(C-C)δ(C–OH) δ(O-CH)  | Cellulose (particularly C3-O3H secondary alcohols) polysaccharidesPectin Secondary metabolites (monoterpenes from leaf) | [1], [3], [5], [6], [9] |
| 996 a | ν(C-O) ring stretching ν(C-C) ring breathing | CelluloseNucleic acids | [1], [5], [7] |
| PC-2 (22.6%) | 1598 a | ν(COO)ν(C-C)ν(C=C)ν(C=N)ν(NH2)  | PectinLigninSecondary metabolites (aromatic compounds)Nucleic acids | [3], [4], [5], [6], [8], [9] |
| 1537 d | Amide II (incl. secondary structure in β-sheet motifs)ν(C=N) δ(NH2) | ProteinsLigninNucleic acids | [1], [3], [5], [7], [8]  |
| 1102 a | ν(C-O-C) ester ν(C-O) ν(C-C) νs(PO42-) | CutinCellulose Pectin (as acetyl-ester)PolysaccharidesDNA/RNA | [1], [2], [4], [5], [6], [8], [9] |
| 1037 d | ν(C-O) ν(C-C)ν(CCO)ν(O-CH3) ν(CH2OH) | CellulosePectin (also as acetyl-ester)PolysaccharidesXyloglucanArabinogalactan Galactan | [3], [5], [8], [9] |
| **945** | ω(CH2)  | PectinProtein phosphorylationSecondary metabolites (monoterpenes from leaf) | [5], [8], [9] |
| **Table References** | [1] Abidi et al. 2014; [2] Butler et al. 2015; [3] Butler et al. 2017; [4] Heredia-Guerrero et al. 2014; [5] Largo-Gosens et al. 2014; [6] Monti et al. 2013; [7] Movasaghi et al. 2008; [8] Ord et al. 2016; [9] Schulz and Baranska 2007 |
| **Bold** wavenumbers represent unique spectral markers for each class. a : wavenumber overlap between all infection categories (PS, IS, and AS); b : wavenumber overlap between infection categories PS and IS only (not observed); c : wavenumber overlap between infection categories PS and AS only; d : wavenumber overlap between infection categories IS and AS only; ν: vibration; δ: deformation; ω: wagging. |

Table 2. Validation set classification results of control versus infected tomato plants using PCA-LDA.

|  |  |  |  |
| --- | --- | --- | --- |
| **Infection Category** | **Accuracy** | **Sensitivity** | **Specificity** |
| Pre-symptomatic | 100% | 100% | 100% |
| Intermediate Symptomatic | 100% | 100% | 100% |
| Advanced Symptomatic | 100% | 100% | 100% |