Plant Science xxx (xxxx) xxxx

Contents lists available at ScienceDirect

Plant Science

journal homepage: www.elsevier.com/locate/plantsci

Investigating potato late blight physiological differences across potato cultivars with spectroscopy and machine learning

Kaitlin M. Gold[a,*], Philip A. Townsend[b], Ittai Herrmann[c], Amanda J. Gevens[d]

[a] University of Wisconsin-Madison, Department of Plant Pathology, United States
[b] University of Wisconsin-Madison, Department of Forestry and Wildlife Ecology, United States
[c] The Robert H. Smith Institute for Plant Sciences and Genetics in Agriculture, The Hebrew University of Jerusalem, P.O. Box 12, Rehovot 7610001, Israel
[d] University of Wisconsin-Madison, Department of Plant Pathology, United States

ARTICLE INFO

Keywords:
Phytophthorainfestans
Reflectance spectroscopy
Hyperspectral
Machine learning
Cultivar discrimination

ABSTRACT

Understanding plant disease resistance is important in the integrated management of *Phytophthora infestans*, causal agent of potato late blight. Advanced field-based methods of disease detection that can identify infection before the onset of visual symptoms would improve management by greatly reducing disease potential and spread as well as improve both the financial and environmental sustainability of potato farms. *In-vivo* foliar spectroscopy offers the capacity to rapidly and non-destructively characterize plant physiological status, which can be used to detect the effects of necrotizing pathogens on plant condition prior to the appearance of visual symptoms. Here, we tested differences in spectral response of four potato cultivars, including two cultivars with a shared genotypic background except for a single copy of a resistance gene, to inoculation with *Phytophthora infestans* clonal lineage US-23 using three statistical approaches: random forest discrimination (RF), partial least squares discrimination analysis (PLS-DA), and normalized difference spectral index (NDSI). We find that cultivar, or plant genotype, has a significant impact on spectral reflectance of plants undergoing *P. infestans* infection. The spectral response of four potato cultivars to infection by *Phytophthora infestans* clonal lineage US-23 was highly variable, yet with important shared characteristics that facilitated discrimination. Early disease physiology was found to be variable across cultivars as well using non-destructively derived PLS-regression trait models. This work lays the foundation to better understand host-pathogen interactions across a variety of genotypic backgrounds, and establishes that host genotype has a significant impact on spectral reflectance, and hence on biochemical and physiological traits, of plants undergoing pathogen infection.

1. Introduction

Understanding plant disease resistance is critical for successful integrated disease management. This is particularly true for potato, where cultivars have high levels of variation in major disease resistance. Late blight of potato, caused by the oomycete pathogen *Phytophthora infestans*, is one of the most devastating diseases to sustainably and effectively control. Considerable efforts have been made to advance disease resistance breeding in conventional programs through the ingression of QTLs from diploid potatoes [1,2] and genetic manipulation [3,4]. Despite these advances, new potato cultivar adoption is a challenging due to phenotypic variation selected for different field, storage, and processing sectors. Growers rely upon chemical control for late blight management [5] due to the lack of effective resistance in commercially desirable cultivars and high potential risk of crop yield and quality loss. Further complicating integrated late blight management, potato cultivars can exhibit variable resistance to late blight across plant organs. Foliar and tuber plant parts can have different levels of resistance, such as the cultivar Umatilla, which has susceptible foliage and resistant tubers. Other cultivars exhibit foliar resistance with little to no tuber resistance [6].

Reflectance spectroscopy has emerged as an effective approach for fast, non-destructive estimation of a wide variety of plant chemical, morphological, and metabolic traits in living tissue [7–17]. Changes in leaf optical properties arise from the interaction of light, chemical bonds, and leaf cell structure. This phenomenon allows us to directly estimate foliar structure, plant chemical composition, water concentration, and metabolic status from reflectance measurements in the visible (VIS), near-infrared (NIR), and shortwave infrared wavelengths (SWIR, collectively 350 − 2500 nm). To do this, field spectrometers, handheld or otherwise portable devices that measure light reflectance continuously across a range of wavelengths, are used. These

* Corresponding author.
E-mail address: kmorey@wisc.edu (K.M. Gold).

https://doi.org/10.1016/j.plantsci.2019.110316
Received 28 June 2019; Received in revised form 10 October 2019; Accepted 14 October 2019

measurements differ from other broad-band devices and cameras by covering a large spectral range (visible to SWIR) at narrower intervals (typicall 3 – 8 nm intervals).

Near infrared spectroscopy has been used since the early 2000s to discriminate between crop cultivars in melon [18], coffee [19], wheat [20], oilseed rape [21], cannabis [22], and tea [23]. One of the first studies to include shortwave infrared (SWIR, 1200 – 2500 nm) wavelengths was Rao et al., who found these wavelengths to be important for discrimination among cultivars of rice, cotton, chili pepper, and sugarcane [24]. This study also noted that cotton cultivars that appear visually similar (in visible wavelengths, 400 – 700 nm) could have vastly different reflectance in other spectral regions.

Plant resistance responses, such as reduced colonization, reduced sporulation, and general reduction in disease severity may be identifiable in spectral reflectance [25]. The phenomenon of tolerant cultivars succumbing to infection yet having a different spectral reaction to infection than a susceptible cultivar has been in a handful of publications across plant and pathogen species [26–28]. Cultivar has also been found to influence the accuracy of abiotic stress classifications. Pre- and post-symptomatic spectral response to apple bruising has been found to be variable across cultivars despite almost identical visual damage [29,30].

In potato production, cultivar identity is currently policed through the seed potato certification process which includes visual inspections and periodic genotypic testing [31]. Cultivar purity requires considerable expertise and has become more difficult in recent years due to increased variety in global commercial production. Different potato cultivars may have minor to moderate physiological differences, such as leaf shape and color, or biochemical differences, all of which may impact spectral reflectance. Reflectance spectroscopy can be used to differentiate between healthy potato cultivars with accuracy ranging from 56 to 90%, depending on cultivar [32]. Couture et. al established that accurate *Potato virus Y* detection was possible regardless of cultivar, but that potato cultivars can have different physiological and biochemical responses to infection [32].

The objectives of this work were to (1) determine the effect of cultivar on spectroscopic late blight detection, (2) quantify the difference in cultivar response to *P. infestans* infection using remotely sensed biochemistry and physiology metrics, (3) identify spectral regions most important for detection and discrimination with three different statistical methods.

2. Materials & methods

2.1. Plant and pathogen materials

'Katahdin,' 'Snowden', 'SP951,' and 'Russet Burbank' potatoes from the Wisconsin Seed Potato Certification Program were grown for four weeks from tissue culture cuttings in pathogen-free growth chambers with a 12 h photoperiod under 24 °C daytime temperatures and 21 °C night temperatures. 'SP951' shares a genetic background with 'Katahdin,' but was transformed to contain a single copy of the Rb late blight resistance gene (from wild potato *Solanum bulbocastanum*) [3]. A representative *P. infestans* isolate of US-23 clonal lineage was collected from an infected potato field in 2017 in Wisconsin, purified, and maintained on Rye-A agar plates for three-weeks at 18 °C. Cultures were placed at room temperature for 24 – 48 h to better induce sporulation prior to inoculation. A 5-mm agar plug was excised from cultures in areas with concentrated sporulation and inoculated onto the adaxial surface of the leaf. Humidity chambers were used to ensure 100 % ambient humidity to support the infection process. Two treatments were used in this study: non-inoculated Rye A agar plug inoculation (agar control) and US-23 inoculation. Ten replicates were used per treatment and cultivar combination and two leaves measured per plant. Disease was rated as 0–5, with 0 = no disease and 5 = severe disease at each measurement using a modified Horsfall-Barratt scale. Infection was confirmed with both visual assessments and an ELISA-based immunostrip test for the genus *Phytophthora* (Agdia, Elkhart, IN). Data used in this analysis were collected over the course of three independent experiments.

2.2. Reflectance measurements

Contact leaf reflectance was measured using a high-spectral-resolution SVC HR-1024i (350–2500 nm) field spectroradiometer (Spectra Vista Corporation, Schnectady, NY, USA). All measurements were taken from the leaf adaxial surface using a leaf-clip assembly attached to a plant probe with a halogen light source, using 99 % spectralon as white reference (Labsphere, North Sutton, NH). Reflectance was measured on two locally-inoculated leaves per plant with two spectra averaged per leaf location. Measurements were taken next to the inoculation zone, but not on top of the inoculation plug. Reflectance curves were interpolated to 1 nm spacing from the native 3 – 8 nm resolution, and reduced to 400–2400 nm by removing the wavelengths with higher relative noise at the edges of spectra [10,11,33]. Baseline measurements were taken immediately prior to inoculation, and then at 24 h intervals for the following 5–7 days, until disease progressed to sporulation. At each time point images were taken with a Nikon digital camera.

2.3. Data preparation

Spectral measurements with anomalous low reflectance or abnormalities due to measurement error were identified manually and removed. Inoculated plants that did not achieve disease rating of 4 or greater by the end of the study period were removed. Disease time was defined categorically for statistical analyses as follows: "early infection"- late blight inoculated leaves at time points within 24 h of inoculation, "biotrophy" - late blight inoculated leaves at time points occurring after 48 h until the last time point measured before disease symptoms appeared, "necrotrophy" - late blight inoculated leaves during time points after which disease symptoms had become visible but before sporulation occurred, and "sporulation" - inoculated leaves during time points where both necrosis and sporulation were visible. Disease time was defined for control plants as the approximate range of time corresponding to the majority of diseased-treated plants progressed into the next infection progression stage. Data were binned so that the interaction of treatment, cultivar, and time could be estimated while accounting for the observation that infection and disease severity did not progress uniformly for all individuals. This allowed for spectral responses caused by measurement and handling over time to be differentiated from those caused by infection. Due to low disease incidence in the first experiment, there was an unequal number of control and infected samples (5X). To reduce bias, 10 control samples per cultivar per disease time point were randomly selected to be used in further analyses resulting in a total n of 1330 spectral measurements.

2.4. Data analysis

2.4.1. Univariate analysis

All analyses were conducted in R. Normalized difference spectral indices (NDSIs, equation 1) were calculated for all possible combinations of wavelengths (~4 million combinations) to identify which combinations best correlated (Pearson correlation) with change in infection status. NDSIs were calculated across all disease times for each cultivar as well as within each disease stage. The NDSI correlations with disease status (infected vs control) were then plotted as heatmaps to visualize important spectral regions for distinguishing infection status and compare responses across cultivars and stages. Wavelengths of the top 0.1 % (~4000) most correlated NDSIs by Pearson correlation for late blight disease stages were identified and their relative frequency of occurrence and plotted. The relative frequency of a wavelength by

cultivar and developmental stage was plotted to visualize spectral disease progression.

2.4.2. Multivariate analysis

Permutational multivariate analysis of variance using distance matrices (PERMANOVA) was conducted to generate "pseudo" F and p-values for which to test significance influence of both treatment and cultivar on spectral reflectance using the R package *vegan* [34]. The PERMANOVA randomization was restricted (blocked) by experiment and stratified by sample ID and disease time to account for the repeated measures nature of the experiment [35]. Principal coordinates analysis (PCoA) was used to visualize broad spectral differences between infected and non-infected samples of the four cultivars in the study using spectral reflectance as the predictor variables using the R package *vegan* [34]. This method uses a dissimilarity matrix calculated using Euclidean distance to transform a large number of possibly correlated predictor variables into a smaller number of uncorrelated latent variables, or principal coordinates, reducing the dimensionality of the data.

Two discrimination methods were then used to classify infection by cultivar using full-spectrum VSWIR data: partial least squares discriminant analysis (PLS-DA) and ensemble random forest classification [36] with the R packages *pls* and *caret* [37,38]. "Early infection" and "biotrophy" disease time points were binned together into a "pre-symptomatic" category and "necrotrophy" and "sporulation" disease time points were binned together into a "post-symptomatic" category to increase the number of observations category for the classification. PLS-DA discriminates groups based using high dimensional and highly collinear data through the projection of latent variables through the response and predictor variables to maximize explanation of the dependent variable as a function of an optimized reduced set of components [39,40]. PLS-DA analyses were performed using a 70:30 permutational approach to estimate cross-validation and uncertainty following Couture et al. and Herrmann et al. [32,41].

We compared PLS-DA results to results of random forest classifications. Random forest (RF) classification is a non-parametric, ensemble machine learning that uses multiple decision trees to classify data and is well suited to spectral data analysis [36,42,43,60]. For random forest classification, the dataset was resampled to a 10 nm resolution (202 wavebands) with the *prospectr* package in R [44] to yield an optimal number of predictor variables and to reduce collinearity between wavelengths. Random forest was applied to first order derivative (FOD) spectra in order to enhance spectral features, reduce systematic differences, and potentially increase classification accuracy [45]. We performed this comparison due to the inherent nature of tree-based models to consider variables sequentially, making them handy for considering interactions without specifying them. In contrast, PLS-DA does not directly include interactions although the nature of PLS-DA components does capture interactions. RF differ from PLS-DA in that the specific weighting and direction of individual independent variables is not readily identifiable, although the relative importance of a variable can be determined through its prominence within the ensemble of trees. Interactions that are useful for prediction, such as cultivar, are easily incorporated with a large enough forest (ensemble). Two random forest classifications were fit, one including cultivar as a predictor variable, and one without to compare the effect of cultivar on classification accuracy.

The two parameters primarily responsible for random forest model performance, the number of randomly selected predictors to choose from at each split (*mtry)* and number of trees generated to yield a full ensemble (*n-tree*) were optimized in an 80:20 training/testing data subset and 10-fold repeated internal cross-validation to estimate out-of-box accuracy. This process was repeated 10 times and the mean accuracy and kappa [46] were calculated in order to gauge model performance. Kappa is a model assessment that can be understood as model accuracy accounting for accuracy due to random chance and takes on a value between 0–1 [46]. Kappa values can be interpreted using standards established in Landis & Koch (1977): 0 = poor, 0.1-0.2 = slight, 0.2-0.4 = fair, 0.4-0.6 = moderate, 0.6-0.8 = substantial, 0.8–1 = near perfect. Final model parameters used were *n-tree* = 500 with *mtry* varying by model. Average test accuracy kappa, overall accuracy, false negative rate, and false positive rate were used to determine model accuracy.

Foliar nitrogen, total phenolics, sugar, starch, and leaf mass per area (LMA) were estimated from spectra using partial least squares (PLS) regression calibrations from Chlus et al. (in preparation), a well established method in functional ecology [10–12]. LMA is an indicator of leaf dry-mass investment and relates to light interception and leaf longevity [47]. Lower LMA values indicate reduced leaf thickness, meaning that the leaf is allocating more resources to photosynthesis or defense compounds than structural compounds. Allocating more resources towards leaf structure (higher LMA) can protect leaves from desiccation or herbivory [47]. The normalized difference water index (NDWI) was calculated using the relative difference in reflectance at wavelengths 857 nm and 1241 nm as an estimate of leaf water concentration [48]. A two-way, repeated measures ANOVA was fit using a mixed effect model with maximum likelihood estimation method for each of the remotely sensed metrics, total phenolics, sugar concentration, starch concentration, nitrogen concentration, NDWI, and LMA using the R package *nmle* [49]. The R package *emmeans* was to elucidate interaction effects for cultivar, treatment, and the interaction of cultivar and treatment across time [50].

3. Results

3.1. Cultivar affects spectral profile across all disease time stages

Disease progressed similarly for all treatments included in this study. No significant differences were identified in area under the disease progress curve (AUDPC) between late blight inoculated cultivars (Fig. S1). AUDPC was not significantly different between 'Katahdin' and 'SP951' cultivars (t-test, p-value 0.71). Principle coordinate analysis (PCoA) showed a strong cultivar effect on spectral reflectance with a high degree of separation between 'Russet Burbank' and 'Snowden' potatoes from 'Katahdin' and 'SP951' potatoes, regardless of infection status (Fig. 1). Visualizing the data as such, the infection effect is masked due the strong cultivar influence (Fig. S2). 'Katahdin' and 'SP951' share a genetic background and showed a high degree of overlap between their respective spectral profiles. As disease stage progressed, inoculated and non-inoculated plants became more different from each other for all four cultivars (Fig. 1). Across all four cultivars, infected and non-infected samples showed separation at all four disease time stages (Fig. 1). The first two principal coordinates explained 90 + % of total of the variation seen in the spectral data.

Cultivar had a strong impact on spectral reflectance but it was difficult to distinguish between different cultivars with visible range reflectance alone. 'Katahdin' and 'SP951' plants consistently had higher reflectance in the NIR than 'Snowden' and 'Russet Burbank' potatoes (Fig. S3). A weakly significant treatment effect (pseudo pvalue = 0.048), a highly significant cultivar effect (pseudo $p < 0.001$), and a weakly significant interaction effect ($p = 0.037$) were seen across all time points in the study (Table S1). PERMANOVA was calculated within each disease time stage to understand effects of disease time on treatment and cultivar interactions. During early infection, a highly significant cultivar effect was seen ($p < 0.001$) but there was no evidence for a significant treatment or interaction effect (Table S1). Significant treatment, cultivar, and interaction effects during biotrophy, necrotrophy, and sporulation were seen. This indicates that the impact of both cultivar and inoculation on vegetation reflectance differs depending on the treatment and cultivar combination.

NDSI heat maps (Fig. 2) were generated to visualize the broad changes in vegetation reflectance caused by infection. Overall, the

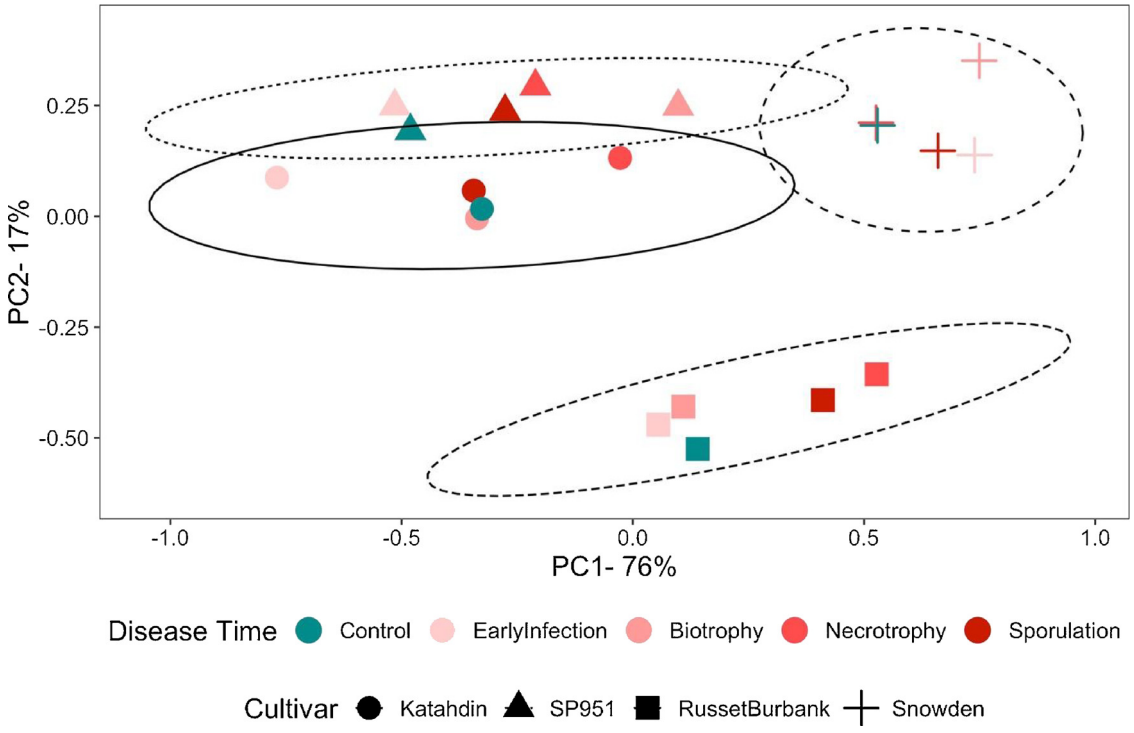

Fig. 1. Principle Coordinate Analysis (PCoA) using euclidean distance showing strong cultivar effect on spectral reflectance regardlss of infection status over disease time. 'Katahdin' and 'SP951' share a genetic background and showed a high degree of overlap between their respective spectral reflectance.

cultivars showed a great deal of variation in which NDSIs showed the greatest change during infection. When all cultivars were compared across all disease time points, few noticeably different regions could be seen (Fig. 2A). When all cultivars were combined and compared across individual disease time stages (Fig. 2F, K, P, U), few highly changed regions can be seen until sporulation (Fig. 2U). Binning together all disease time points, the four cultivars showed different spectral responses to inoculation (Fig. 2B–E). When fit across all disease time points, the wavelength combinations most strongly associated with infection in 'Katahdin' plants utilized visible bands in conjunction with all areas of the spectra (VNIR through SWIR), as well as regions in the SWIR (Fig. 2B). Fine narrow band regions of the NIR also showed a difference in reflectance. 'SP951' plants, which share a genetic background with 'Katahdin' plants, also differed in SWIR wavelengths > 2000 nm but differed from 'Katahdin' plants in showing a much higher associations with NIR wavelengths and a lesser impact of VIS wavelength interactions with the full spectra (Fig. 2C). Overall, 'Katahdin' and 'SP951' plants appeared to have similar spectral responses to infection, but infection caused stronger change in the NIR wavelengths of 'SP951' than 'Katahdin'.

Infection in 'Russet Burbank' plants showed a significant correlations between infection status and NDSIs using SWIR bands, but overall, had a much broader influence over in the NIR, especially NDSIs using longer NIR wavelengths with SWIR and VIS wavelengths. Fine narrow band features in the NIR and VIS differed considerably by infection (Fig. 2D). In contrast, infection led to differences at very narrow wavelengths in 'Snowden' plants, with a specific effect on narrow band NIR-vs-NIR interactions (Fig. 2E). A weak effect was seen in interactions of the SWIR wavelengths in the 2000s that was shared amongst the cultivars.

NDSIs were fit for each disease time point for all cultivars combined to identify whether there were general responses across cultivars (Fig. 2, bolded box). During early infection (Fig. 2G–J) and biotrophy (Fig. 2L–O), SWIR narrowband features were most strongly affected by infection in all cultivars. These narrow band features are likely associated with changes in water content and potentially nitrogen concentration and structural carbohydrates (Serbin et al., 2012, 2014). During early infection, all four cultivars showed a shared response in the narrowband features of the SWIR wavelengths, with greatest overlap between 'Katahdin' and 'SP951' (Fig. 2G–J). 'SP951' plants showed a greater reaction in the interactions of the NIR wavelengths than 'Katahdin' plants that was shared with 'Russet Burbank' and 'Snowden' plants. During this stage, SWIR features in the wavelength region approaching 2400 nm were most important, with a strong feature developing in the NIR. During sporulation, the same strong SWIR feature distinguished necrotrophy from non-inoculated control leaves, along with a strong impact in visible wavelengths, as would be expected.

3.2. Cultivars differ in remotely sensed biochemical and physiological indices at different disease time stages

To gain a better understanding of how cultivars respond to early stage infection, spectrally-derived biochemical and physiological indices, derived via PLS-regression, were calculated and compared. Overall, disease time had a highly significant effect on all non-destructively sensed metrics except for NDWI where no evidence of a significant effect was seen (Table S2, Fig. S4). There was a highly significant cultivar effect on concentration of total phenolics, sugars, LMA, and NDWI, but no evidence of a significant effect on starches and nitrogen concentration (Table S2). Treatment had a significant effect on phenolics and starch concentration. When taking into account disease time, a significant treatment and disease time interaction was seen for phenolics, starches, nitrogen, LMA, NDWI, and a weakly significant effect on sugars. A moderately significant three-way interaction of disease time, cultivar, and treatment was seen on NDWI and weakly significant for sugars and LMA, indicating that these traits changed across cultivar and treatment with disease time (Table S2).

Pairwise comparisons accounting for the random effect of disease time for treatment over levels of disease time and cultivar (Table S3, annotated on Fig. S4) and for cultivar over levels of disease time and treatment (Table S4) were performed to better understand how the

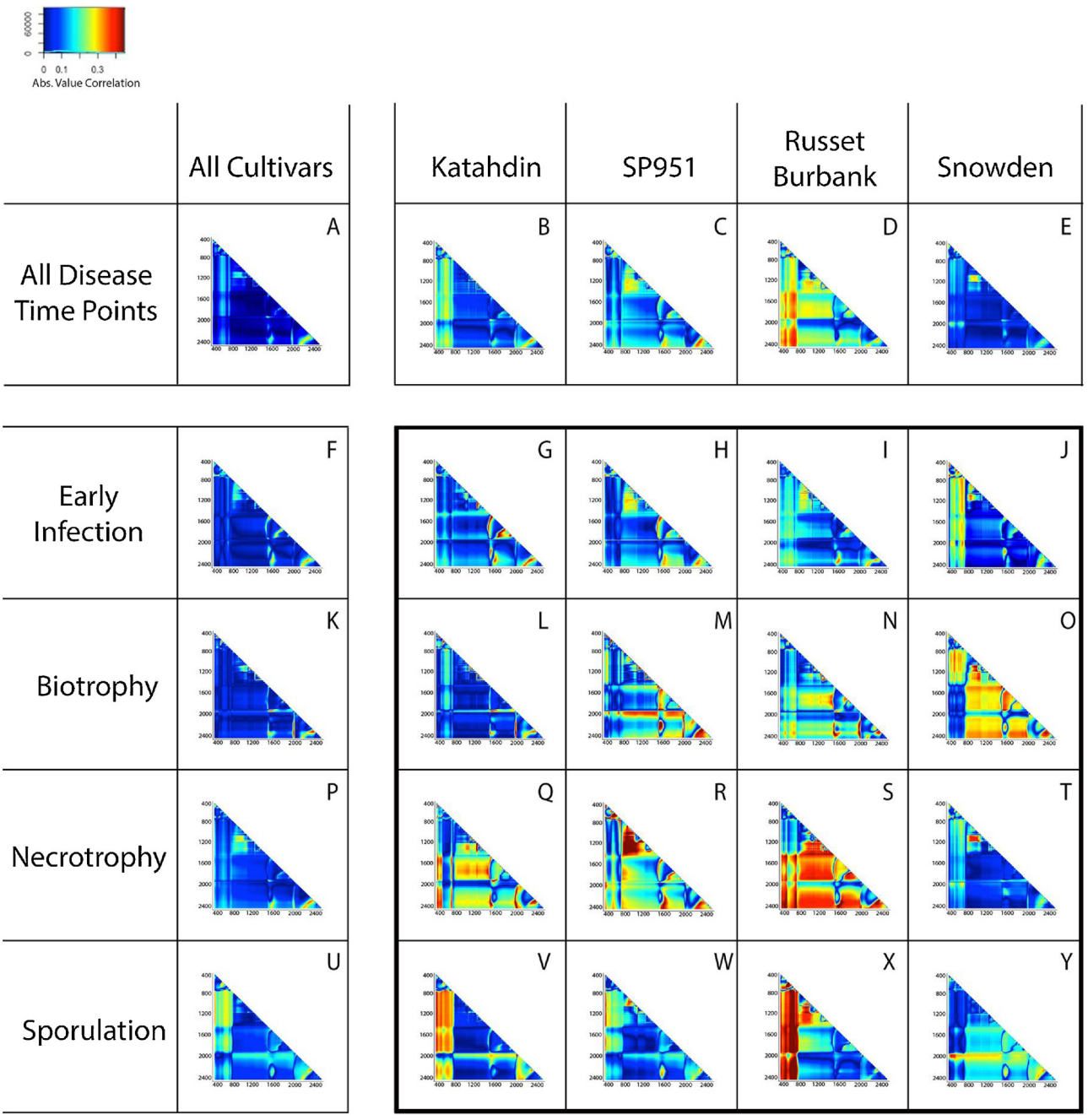

Fig. 2. Normalized difference spectral index (NDSI) heat maps across **(A)** all disease time stages and cultivars, **(B–E)** individual cultivars across all disease time stages, **(F, K, P, U)** all cultivars across individual disease time stages, **(G–J)** individual cultivars during early infection, **(L–O)** individual cultivars during biotrophy, **(Q–T)** individual cultivars during necrotrophy, **(V–Y)** individual cultivars during sporulation. Red values indicate high absolute value of Pearson's correlation with a change in infection status (non-inoculated vs. inoculated).

differences and similarities in how cultivars respond to disease. For the cultivar pairwise comparisons, only comparisons that had a significant difference between infected leaves without a significant difference between the pairwise comparison of control leaves (or vice versa) were considered to be noteworthy. Cultivars likely have different base concentrations of phenolics, and the stress of time course measurement could potentially change values, which is why it is important to consider one pairwise comparison (such as infected 'Katahdin' vs infected 'SP951') in light of the alternate pairwise comparison (control 'Katahdin' vs control 'SP951').

Phenolics concentration differed significantly between infected and healthy 'Katahdin' leaves during early infection, but not for 'SP951' or any of the other cultivars (Table S3). During this disease time stage, 'Katahdin' and 'SP951' infected leaves had a significantly different phenolics concentration (pvalue = 0.01) from each other but the control leaves did not (pvalue = 0.28, Table S4). Infected 'SP951' leaves had a significantly different leaf sugar concentration than infected 'Snowden' leaves (pvalue = 0.011) while a pairwise comparison of control leaves yielded no significant difference between them (pvalue = 0.029, Table S4). Sugar concentration was weakly significantly different between infected and control leaves for both 'Katahdin' and 'SP951' leaves (Table S3). There was no difference in starch concentration between treatments (control and infected leaves) during early infection for any of the four cultivars (Table S3), but infected 'Russet Burbank' leaves had significantly different concentration of phenolics from both 'Katahdin' and 'Snowden' leaves (pvalues 0.024

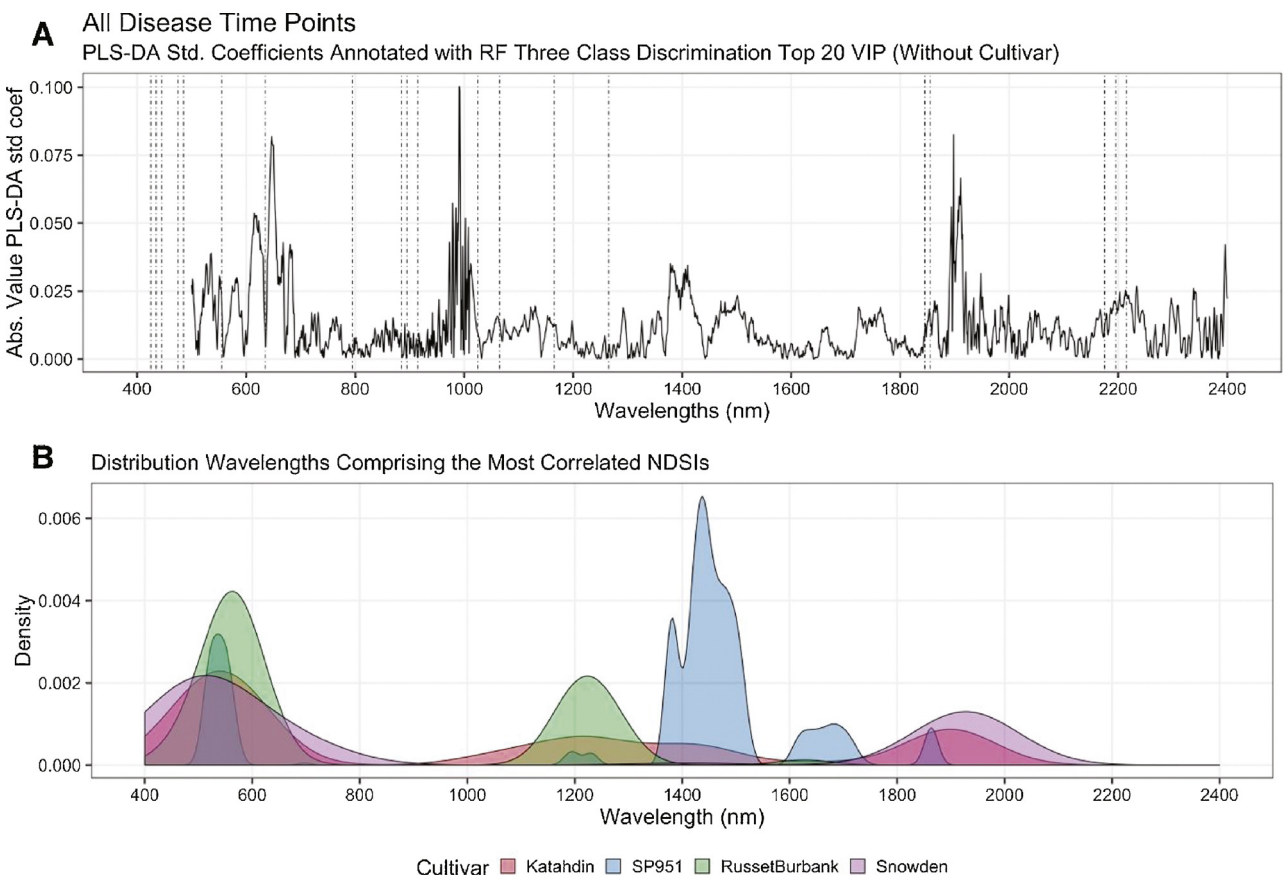

Fig. 3. A) PLS-DA absolute value of standardized coefficients by wavelength overlaid with top 20 VIPs from three class RF discrimination (without cultivar) **B)** Distribution of top 0.1 % of most correlated NDSI wavelengths from all disease time points.

and 0.044 respectively) while there was no significant difference in control leaves (Table S4). 'Russet Burbank' infected leaves had significantly different nitrogen concentration than control 'Russet Burbank' leaves (pvalue = 0.025), infected 'SP951' leaves (pvalue = 0.011), and infected 'Snowden' leaves (pvalue = 0.006, Table S3, S4). LMA and NDWI were weakly significantly different between infected and control 'Russet Burbank' leaves as well (pvalues 0.047 and 0.0608 respectively). NDWI was significantly different between infected 'SP951' leaves and both control 'SP951' leaves (pvalue = 0.0436) and infected 'Katahdin' leaves (pvalue = 0.011).

During biotrophy, 'Katahdin', 'SP951', and 'Snowden' had highly significant differences in phenolics concentration between infected and control leaves (Table S3). Interestingly there were highly significant differences between 'Russet Burbank' control leaves and control 'Katahdin' (pvalue = 0.0005), and 'SP951' (pvalue < 0.001), leaves, that were not seen in pairwise comparison of respective infected leaves (Table S4). A similar pattern was seen with LMA, in which the pairwise comparison between control 'Russet Burbank' leaves and both control 'Katahdin' (pvalue = 0.009) and control 'SP951' (pvalue = 0.0013) leaves. 'Katahdin' and 'Snowden' control leaves were significantly different from each other as well (pvalue = 0.0019), but the pairwise comparison of their respectively infected leaves was not (pvalue = 0.09). The only other noteworthy trait comparison during biotrophy was the comparison between 'Russet Burbank' infected and control leaves, which yielded significantly different starch, nitrogen, and LMA (Table S3). Infected 'Russet Burbank' leaves had significantly different nitrogen concentration than infected 'Katahdin' (pvalue = 0.022) and 'Snowden' (pvalue = 0.001) leaves (Table S4).

'SP951' infected leaves had significantly different total phenolics, sugar, starch, and NDWI from non-infected leaves during necrotrophy but 'Katahdin' did not. Infected and control 'Katahdin' leaves had significantly different LMA from each other (pvalue = 0.017, Table S4). Control 'Katahdin' leaves had significantly different phenolics concentration from control 'SP951' (pvalue = 0.0395) and control 'Snowden' (pvalue = 0.001) leaves but there was no significant difference between respectively infected leaf comparisons. The same pattern was seen regarding LMA between control 'Katahdin' and 'Snowden' leaves (pvalue < 0.0001). Sugar concentration was significantly different between infected 'Snowden' leaves and infected 'SP951' leaves (pvalue = 0.017) but not in a comparison of their respective control leaves (pvalue = 0.88). Starch concentration was significantly different between control 'Russet Burbank' leaves and control 'SP951' leaves (pvalue = 0.03) but not in a comparison of their respective infected leaves (pvalue = 0.91). NDWI was highly significantly different between infected 'Katahdin' and 'SP951' leaves (pvalue = 0.0004) but not between control leaves (pvalue = 0.2388). Surprisingly, sporulation yielded the least number of significant differences between infected and control leaves amongst the characterized traits. Infected 'Russet Burbank' leaves had weakly significantly different phenolics concentration and NDWI from control leaves (Table S3). Both 'Katahdin' and 'SP951' showed weakly significant differences between infected and control leaves. Infected 'Katahdin' leaves had significantly different sugar concentration from infected 'SP951' leaves during sporulation but not between their control leaves (pvalue = 0.023, Table S4).

Pairwise comparison of biochemical trait concentration and physiological metrics (accounting for interactions of cultivar and disease time) showed some variation between 'Katahdin' and 'SP951' plants across the different disease time stages (Fig. S4, Table S2). Phenolics differed between non-inoculated control and infected 'Katahdin' leaves, but not between control and infected 'SP951' leaves during early infection. During necrotrophy, the opposite was true, with phenolics concentration differing between 'SP951' control and infected leaves but

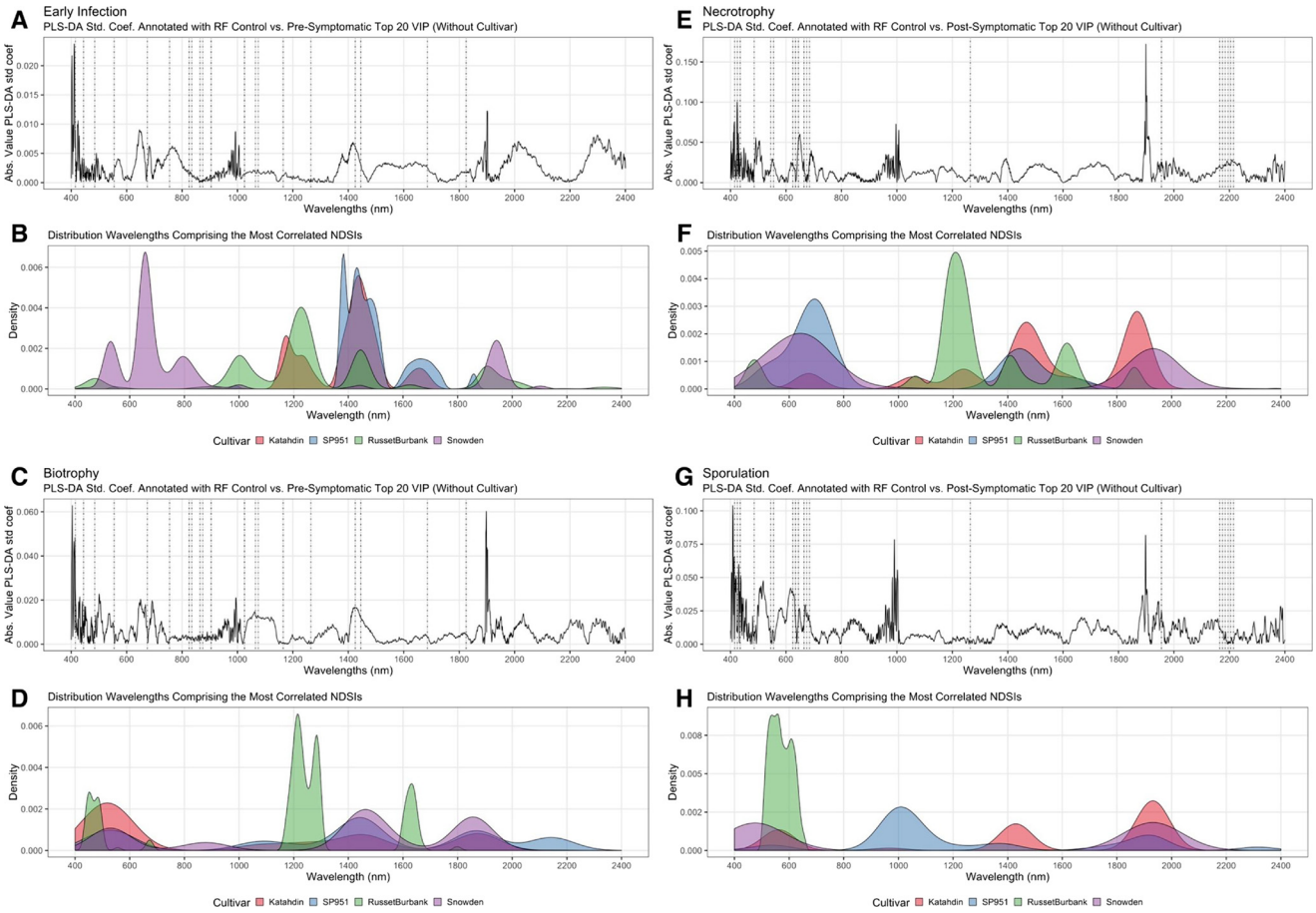

Fig. 4. PLS-DA absolute value of standardized coefficients by wavelength overlaid with top 20 VIPs from RF (A, C, E, G) with density plots of wavelengths comprising the top 0.1 % of most correlated NDSI (B, D, F, H): **A–B)** Early Infection **C–D)** Biotrophy **E–F)** Necrotrophy **G–H)** Sporulation.

not 'Katahdin' leaves (Table S4). Both were significantly different during biotrophy and neither were significantly different during sporulation. During biotrophy, infected 'Katahdin' and 'SP951' leaves had significantly different phenolics concentrations (pvalue = 0.01) from each other but the non-infected control leaves did not (pvalue = 0.28, Table S4). Sugar concentration was weakly significantly different between control and infected leaves of both 'Katahdin' (pvalue = 0.0514) and 'SP951' (pvalue = 0.0327), respectively, during early infection, but only between 'SP951' control and infected leaves during necrotrophy (pvalue 0.001). Sugar concentration of infected 'Katahdin' and 'SP951' leaves was significantly different, but there was no significant difference when compared to non-infected leaves. Starch concentration was significantly different between the non-infected control and infected 'SP951' leaves during necrotrophy and sporulation, but not between 'Katahdin' leaves. NDWI was not significantly different between 'Katahdin' control and infected leaves during any disease time stage but was for 'SP951' control and infected leaves during early infection and necrotrophy. NDWI was highly significantly different between infected 'Katahdin' and infected 'SP951' leaves (pvalue = 0.0004) but not between control 'Katahdin' and 'SP951' leaves (pvalue = 0.2388). LMA was significantly different between control and infected Katadhin leaves during necrotrophy but not for 'SP951'.

3.3. Identifying variation and consistency in important spectral regions across cultivars and disease time

PLS-DA was performed to test the accuracy of classification of infected and non-infected leaves (Figs. 3 and 4, Table 1). Across all disease time stages, the PLS-DA model used 14 components and achieved discrimination accuracy of 70 % and kappa of 0.3. Accuracy was higher when disease progression was taken into account (Table 1). Taking the first order derivative of spectra to reduce systematic variation between cultivars did not increase the accuracy or kappa of the resulting model (data not shown). During early infection, infected and non-infected samples could be discriminated with 65 % accuracy (kappa = 0.214). Discrimination accuracy was 73 % during biotrophic growth (kappa = 0.35), necrotrophic growth (kappa = 0.39) and sporulation (kappa = 0.42) respectively. Binning together the pre-symptomatic disease time stages, discrimination accuracy was 67 % (kappa = 0.19) and 71 % (kappa = 0.39) during post-symptomatic infection (Table 1).

Random forest classification was implemented on first derivative spectral data, both with and without cultivar included as a predictor variable, to compare to PLS-DA and determine whether a machine learning method could yield higher accuracy and greater model robustness. Random forest discrimination had higher accuracy than PLS-DA in detecting infected leaves across cultivars, but both identified similar spectral features (Figs. 3,4). Accuracy and kappa metrics between random forest classification and PLS-DA were comparable (Tables 1 and 2). Random forest without cultivar as a predictor could classify between control, pre-symptomatically, and post-symptomatically infected plants with ~70 % accuracy (kappa = 0.3). When cultivar was included, accuracy did not increase significantly (71 %) but kappa increased to 0.35. With and without cultivar included, control and pre-symptomatic infected plants could be distinguished with 83 % accuracy (kappa = 0.42) and 85 % accuracy (kappa = 0.45) respectively. Control and post-symptomatic infected plants could be distinguished with 83 % accuracy (kappa = 0.37) when cultivar was included and 82 % accuracy (kappa = 0.42) without cultivar as a predictor variable. Pre-

Table 1
PLS-DA discrimination accuracy across all disease time stages and within each disease time stage. Internal cross validation performed with 100 permutations with a 70/30 calibration-validation data set split.

Partial Least Squares Discriminant Anaylsis (PLS-DA)						
Model	Train		Internal Cross-Validation		Components	Top 20 Standardized Coefficients (nm)
	Accuracy	Kappa	Accuracy	Kappa		
All Disease Time Points	72.23%	0.37	68.60%	0.29	14	1898, 1911, 1908, 1909, 1912, 1910, 1894,1906, 615, 619, 617, 618, 616, 623, 1907,988, 614, 624, 1008, 620
Early Infection	71.02%	0.34	65.09%	0.21	7	407, 429, 1894, 1893, 2013, 406, 2016,2015, 2014, 2012, 2000, 2021, 2017, 2018, 2019, 2001, 2020, 2022, 2002, 2011
Biotrophy	78.65%	0.49	73.16%	0.35	10	1898, 1897, 1896, 410, 1900, 1901, 1902, 411, 409, 412, 501, 502, 431, 995, 1908, 994, 692, 500, 505, 1909
Necrotrophy	83.10%	0.62	72.73%	0.39	13	1899, 1902, 1900, 1901, 1895, 414, 997, 413, 1007, 428, 1908, 1906, 1904, 1910,491, 1905, 1903, 1907, 1909, 1911
Sporulation	80.68%	0.61	71.30%	0.42	12	408, 409, 1898, 413, 410, 430, 412, 429, 437, 438, 436, 617 616, 1899, 615, 618, 619, 623, 624, 614

and post-symptomatic infected plants could be distinguished from each other with 71 % accuracy (kappa = 0.41) and 69 % accuracy (kappa = 0.37) with and without cultivar included as a predictor variable respectively. Overall, high variability between cultivars may reduce model accuracy, but we found that including cultivar as a potential predictor variable did not dramatically change model accuracies and kappas. Three-class discrimination between control, pre-symptomatic, and post-symptomatic samples was performed for each cultivar (Table 3). Model performance varied considerably between cultivars. 'SP951' and 'Snowden' discrimination accuracy was poor (kappa = 0.18 and 0.14 respectively) whereas 'Katahdin' and 'Russet Burbank' discrimination accuracy performed moderately well in comparison (kappa = 0.4 and 0.37 respectively).

Examination of the wavelengths used to make the top 0.1 % of NDSI (approximately 4000 wavelengths), as ranked by Pearson's correlation for each cultivar as infection progressed, confirmed our previous finding that important regions of wavelengths change as disease time progresses [51 in review]. However, we found that the specific spectral regions that are most important during each stage is cultivar dependent (Fig. 4B, D, F, H). During early infection in 'Katahdin' and 'SP951' plants, regions centered around the water absorption feature at 1400 nm are the most divergent from non-inoculated control plants. In 'Katahdin' plants, the most highly correlated regions are sharply centered around these water bands, but inoculated 'SP951' foliage showed a reaction to infection at this stage across a broader range of wavelengths spanning the 1400 nm water band through the 1700 nm (Fig. 4B). In contrast, visible wavelengths were most strongly affected by infection in 'Snowden' plants and the red edge (700 − 750 nm) and 1000 nm NIR plateau in 'Russet Burbank' plants (Fig. 2). During biotrophic growth (Fig. 4D), 'Katahdin' plants showed a strong spectral response in the visible range, 'SP951' plants exhibit the greatest change in the 1400 and 1900 nm water features, 'Russet Burbank' plants show the strongest response in the NIR plateau, and 'Snowden' plants were most strongly affected in the visible and SWIR water band range (Fig. 4F, 5H).

During necrotrophic growth, the most highly correlated wavelengths for 'Katahdin' were centered around the visible and red edge, as well as a wide range of wavelengths spanning the SWIR between 1400 − 2000 nm. In contrast, the most highly correlated wavelengths for 'SP951' were solely in the visible range (specifically centered on a small range of wavelengths in the 600 nm region). The most highly correlated wavelengths for 'Russet Burbank' remained in the red edge and NIR plateau (as with biotrophy); 'Snowden' continued to show a strong response in the visible range with the influence in the SWIR shifting to be more centered around the 1890 − 2000 nm during necrotrophy and continuing on through sporulation. During sporulation, 'Russet Burbank' showed the strongest change in the visible range, whereas 'Katahdin' showed the strongest response around 1890 nm. 'SP951' plants showed the strongest change in the NIR bands centered around 1000 nm.

3.4. Three theoretically different analytical methods identify similar important spectral features

All three statistical methods– PLS-DA, RF, and NDSI correlation– identified similar spectral features important for late blight detection across and within cultivars (Figs. 3,4). Strong features associated with the four cultivars were identified by examining the most correlated NDSIs and evaluating random forest VIP and PLS-DA standardized coefficients. Across all disease time points, wavelengths in the visible range were most differentiating for 'Snowden', which can be potentially correlated with the RF VIP and PLS-DA standard coefficient peak around 580 nm (Fig. 4). Random forest VIP and PLS-DA standardized coefficients cannot be directly compared because of different underlying wavelength bin sizes. During early infection, the water absorption feature around 1400 nm was particularly differentiating for 'Katahdin'

Table 2
Cultivar three class (control, pre-symptomatic, and post-symptomatic) random forest discrimination accuracy. Internal cross validation yielding accuracy and kappa values was performed 10 times with data split 80/20 calibration/validation.

Random Forest Three Class Discrimination: First Order Derivative Spectra				
Cultivars	CV Accuracy	CV Kappa	*mtry*	Top 20 VIP (nm)
Control vs Pre-Symptomatic vs Post-Symptomatic				
Katadin	71.13%	0.40	36	485, 1165, 1005, 1205, 1265, 1115, 2205, 685, 475, 1935, 445, 545, 435, 2195, 425, 635, 665, 495, 1285, 1215
SP951	68.82%	0.18	26	1265, 415, 555, 545, 865, 2165, 1035, 2175, 425, 895, 2205, 1025, 875, 1165, 1185, 1435, 1115, 2185, 1045, 905
Russet Burbank	65.98%	0.37	16	2135, 2145, 2075, 475, 745, 2085, 625, 885, 645, 2195, 665, 2175, 2205, 2095, 915, 2185, 485, 635, 715, 725
Snowden	61.68%	0.14	11	445, 1265, 875, 2385, 1115, 1235, 1025, 1095, 1945, 2165, 1085, 1625, 1155, 1245, 1205, 1975, 1715, 2375, 1145, 1955

plants. This is reflected in a PLS-DA standard coefficient peak as well as two RF VIP (Fig. 4A, B). During biotrophy, two RF VIP, 1165 nm and 1265 nm from the control vs. pre-symptomatic differentiation model were aligned with a bimodal peak in the 'Russet Burbank' density plot (green), and can thus be assumed to be highly important for differentiating 'Russet Burbank' healthy and control leaves (Fig. 4C, D).

Despite some variation, both PLS-DA and Random Forest identified similar spectral features for discrimination during all disease time points (Fig. 3A) and within the various disease time stages (Fig. 4). It is important to note that we binned wavelengths to 10 nm intervals for Random Forest classification (necessary because RF is sensitive to number of independent variables), so each identified VIP can account for the influence of nearby bands when comparing to PLS-DA. When compared to the distribution of most correlated NDSI for each cultivar (Figs. 3B, 4 B, 4 D, 4 F, 4 H), the VIP and most correlated standard coefficients can be related to the influence of the underlying cultivars. Where we cannot align a specific cultivar's influence with any one RF or PLS-DA feature, we can assume that the region is important for discrimination across all cultivars. This is the case with VIP in the SWIR 2100 nm range during symptomatic disease time stages (Fig. 4E, G). Together, the three analytical methods elucidate the most important regions for discrimination, and thus, each provides information needed for the long-term goal of accurate, pre-symptomatic field detection, and a deeper understanding of the underlying processes driving our ability to discriminate between infected and healthy plants across cultivars using reflectance spectroscopy.

4. Discussion

The spectral response, and thus the underpinning biochemical and physiological traits, of four potato cultivars to inoculation with *Phytophthora infestans* clonal lineage US-23 was highly variable. There was a significant cultivar by treatment interaction across all timepoints and within each disease time stage except for early infection (Table 1). Thus, we find that cultivar, or plant genotype, has a significant impact on spectral reflectance of plants undergoing *P. infestans* infection. Plant genotype can influence trait expression, and subsequent detection and discrimination [32] especially during early stage late blight progression [52]. Here, we find there to be strong variation in plant response to pathogen infection at early stages of disease (1–6 days post inoculation). At later stages of disease development and progression, there is greater similarity across cultivars. These results serve to emphasize the importance of including multiple cultivars in early disease detection studies and model development due to the high potential for variation in trait response as a consequence of breeding [32]. The inclusion of multiple cultivars is critical in future remote sensing studies, as the most important spectral differentiators can be quite different despite seemingly identical visible disease progression.

We found that late blight changed leaf physiology consistently and towards a more uniform state across all cultivars, including 'Katadhin' and 'SP951'. 'These cultivars share a genetic background, varying only in a single copy of the Rb gene, which confers late blight resistance from the wild type potato *Solanum bulbocastanum* [3]. In our study, while 'SP951' plants exhibited some resistance to late blight, perhaps more aptly termed tolerance, they were not immune. Late blight developed under high inoculum levels and conducive environmental conditions. We did not see a significant difference in AUDPC between 'Katahdin' and 'SP951' plants (Fig. S1B) despite a difference in mean AUDPC. The spectral responses of late blight-infected 'SP951' and 'Katahdin' plants clustered together in our PCoA analysis (Fig. 1), but showed differences in spectral responses to disease at all stages as illustrated by the normalized differential spectral index fit for each cultivar (Fig. 2) and non-destructive assessment of physiological and biochemical status (Fig. S4, Table S4).

Overall, the greatest differences between 'Katahdin' and 'SP951' plants were seen during the earliest stages of infection, during the first three disease time stages (early infection, biotrophy, and necrotrophy). This suggests that the underlying infection processes driving spectral differences during early disease establishment were the most divergent between 'Katahdin' and 'SP951' at these stages. The RB gene responds to the IPI-O family of effectors and may be influencing plant response during early infection, despite the fact that it was overpowered and the leaves ultimately succumbed to disease in time [53]. Despite the fact that 'SP951' succumbed to infection pressure, our work corroborates that the resistance conferred by the RB gene impacts the mechanisms by which the plants physiologically respond to infection during the earliest infection stages, as shown through leaf spectral characteristics.

Despite the strong influence of cultivar on spectral profile, we found that the inclusion of cultivar as a potential predictor variable only slightly increased random forest model discrimination accuracy despite the spectral differences seen between cultivars (Table 3; Fig. 2). However, there was enough similarity across infected leaves of the different cultivars to facilitate accurate detection of infected plants with two different multivariate discrimination methods (Table 1,2). This indicates that despite large spectral variation at times, there is still enough commonality to the infection process across cultivars to allow for accurate, broad detection. This work supports the finding in Thomas et al., which found that annotating a single cultivar's disease response was sufficient to train a support vector machine to detect powdery mildew on six barley cultivars with a high throughput hyperspectral imaging system at later stages of infection [54]. Future work should seek to evaluate whether this finding 1) scales to canopy level deployment (e.g. from drones, planes, or spacecraft) and 2) whether there are conditions, either environmental or pathogen, that may detract from this commonality enough to lead to false negatives.

All cultivars in our study showed individual spectral responses to infection in the visible range, near infrared, and shortwave infrared across all stages of infection, though there was variation in the regions that showed the greatest response to infection across cultivars. Response patterns appeared to be most consistent in the SWIR, whereas VIS and NIR patterns appeared more variable across cultivars (Fig. 2, Fig. 4). While SWIR wavelengths did not show the strongest relationships, they were the most consistent, and therefore appear to be important to overall detectability and repeatability. This consistency may

Table 3
Random Forest discrimination accuracy using first order derivative transformed spectra both with cultivar included as a potential predictor variable and without cultivar as a predictor variable. Internal cross validation yielding accuracy and kappa values was performed 10 times with data split 80/20 calibration/validation.

Random Forest Model Metrics: First Order Derivative Spectra				
Model	Internal Cross-Validation		*mtry*	Top 20 VIP (nm)
	Accuracy	Kappa		
With Cultivar				
Control vs Pre-Symptomatic vs Post-Symptomatic	70.94%	0.35	41	1265, 895, 445, 435, 835, 485, 545, 2175, 635, 795, 2215, 1065, 885, 875, 425, 805, 555, 1165, 855, 865
Control vs Pre-Symptomatic	83.41%	0.37	26	1685, 1075, 1265, 1695, 1065, 865, 2265, 1025, 835, 675, 2175, 905, 875, 1165, 2235, 445, 825, 485, 1435, 455
Control vs. Post-Symptomatic	82.65%	0.42	26	2175, 2185, 2195, 545, 635, 625, 415, 665, 2205, 675, 1265, 2215, 435, 585, 555, 485, 2165, 595, 525, 535
Pre-Symptomatic vs Post-Symptomatic	71.00%	0.41	36	2195,1685,675,635,2185,2275,685,425,2205,2265, 645,2255,2245,2215,535,595,825, 1065,975,485
Without Cultivar				
Control vs Pre-Symptomatic vs Post-Symptomatic	68.68%	0.30	26	445, 1265, 485, 555, 2175, 895, 2195, 635, 795, 1165, 2215, 1025, 435, 475, 885, 1065, 425, 915, 1845, 1855
Control vs Pre-Symptomatic	84.83%	0.45	36	835,1825,1445,675,865,1025,875,555,1075,1165, 835,1825,1445,675,865,1025,875,555,1075,1165
Control vs. Post-Symptomatic	81.74%	0.42	26	2175, 1265, 625, 545, 2195, 2205, 2185, 635, 485, 2215,435, 665, 555, 685, 675, 645, 1955, 2165, 415, 425'
Pre-Symptomatic vs Post-Symptomatic	69.00%	0.37	16	2275, 1685, 2265, 635, 2195, 2185, 685, 2215, 675, 2245, 2175, 2255, 2205, 645, 665, 425, 625, 525, 825, 1695

imply models incorporating SWIR wavelengths could potentially be more robust to unknown or variable environmental and/or genetic differences (e.g., cases where models for genotypes do not exist). SWIR wavelengths have been found to be important for pre-symptomatic *P. infestans* detection and discrimination from *Alternaria solani* [51] and discrimination between *P. infestans* clonal lineages [55].

We found a significant three-way interaction of disease time, cultivar, and treatment on NDWI, and a weakly significant effect on sugar concentration and LMA. “Watersoaking” is a phenomenon associated with late blight infection where cellular structure collapses due to pathogen infection [25,56]. The differences seen in NDWI indicated that the plants recognized the pathogen invasion and were potentially investing energy in maintaining water concentration to fight disease. Alternatively, the pathogen could have triggered the plant to take up more water at the expense of other activity. We found in our previous work that late blight infection changes the total phenolics concentration and sugar concentration of leaves [51,55]. Variability in cultivar sugar concentration and the manner in which infection impacts sugar concentration of leaves may influence accuracy of detection.

Spectral signatures associated with infection in SWIR narrow bands were most similar between cultivars. NDSI plots fit for all cultivars show that there is consistency to potato plant response to late blight infection, particularly in the SWIR and around 1000 nm (Fig. 2). Shared impact of infection on SWIR narrow bands could indicate a conserved biochemical reaction or metabolic shift during infection. Deriving and quantifying the underlying driver of this conserved feature could improve detection accuracy across a broader group of plant genotypes and environmental conditions. We saw similarity in the SWIR for both late blight-infected leaves as well as early blight (causal agent *Alternaria solani*)-infected leaves [51]. Couture et al. reported on a similar SWIR response to *Potato Virus Y* infection across multiple potato cultivars [32], suggesting an even greater similarity to the response of potato to infecting pathogens. The characterization of SWIR responses to infection along with the underlying biochemical, physiological, and morphological changes are important areas ripe for exploration. Exploring quantification of compounds associated with conserved defense pathways may allow us to both better quantify broad plant defense response and develop a warning system for general defense activation.

Non-destructive characterization of biochemical traits enables the non-invasive assessment of host-pathogen interactions and their progress or status. We demonstrate that responses in the SWIR spectral region of potato foliage are conserved among cultivars in response to *P. infestans* infection. SWIR wavelengths in these conserved regions are known to be associated with changing biochemical concentration [7]. Despite both a cultivar and treatment effect, our work demonstrated that there was not a two-way cultivar by treatment or three-way disease time by cultivar by treatment effect for concentration of phenolics. The lack of a significant cultivar by treatment interaction term provided evidence that there may be consistency to the impact of infection on total phenolics concentration across cultivars. This finding suggests that phenolics are an attractive group to query for more specific and robust late blight detection across cultivars. Phenolic compounds have long been understood to play a role in plant defense, with many defense-related genes associated with late blight resistance encoding phenolics pathway constituents [57,58]. For example the PAL-1 gene encodes a lyase that controls the production of phenolics compounds is differentially expressed during infection by different cultivars [58]. Secondary metabolites with phenolic structures are differentially produced during early stages of infection [59]. Remote quantification of secondary metabolites up or downregulated during early infection may be a way to increase detection accuracy regardless of plant genotype.

5. Conclusion

We found that the spectral responses of four potato cultivars to infection by *Phytophthora infestans* clonal lineage US-23 were highly

variable, yet with important shared features that facilitated discrimination. Our work lays the foundation to better understand host-pathogen interactions across a variety of genotypic backgrounds, and establishes that host genotype has a significant impact on spectral reflectance, and hence on the underlying biochemical and physiological traits that facilitate detection, in plants undergoing pathogen infection.

Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgements

We acknowledge funding from the University of Wisconsin WARF Accelerator Program, WI Dept. of Agriculture, Trade and Consume Protection Specialty Crop Block Grant Program, and the UW-Madison CALS Endowed Potato Fellowship supported by Wisconsin potato growers through the University. Additional support was provided by Hatch grant WIS01874 and NASA Advanced Information Systems Technology (AIST) grant 80NSSC17K0244. Special thanks to Eric Larson, John Hammel, Tina Wu, Jaime Spychalla, Haley Knight, Erin Wagner, and Adam Chlus for their assistance with this work. Thank you to Rene Heim for assistance and advice in performing random forest discrimination.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.plantsci.2019.110316.

References

[1] E. Zimnoch-Guzowska, R. Lebecka, A. Kryszczuk, U. Maciejewska, A. Szczerbakowa, B. Wielgat, Resistance to Phytophthora infestans in somatic hybrids of Solanum nigrum L. and diploid potato, Theor. Appl. Genet. 107 (2003) 43–48, https://doi.org/10.1007/s00122-003-1221-4.

[2] S. Jansky, D.I. Rouse, P.J. Kauth, Inheritance of resistance to verticillium dahliae in diploid interspecific potato hybrids, Plant Dis. 88 (2004) 1075–1078, https://doi.org/10.1094/PDIS.2004.88.10.1075.

[3] J. Song, J.M. Bradeen, S.K. Naess, J.A. Raasch, S.M. Wielgus, G.T. Haberlach, J. Liu, H. Kuang, S. Austin-Phillips, C.R. Buell, et al., Gene RB cloned from Solanum bulbocastanum confers broad spectrum resistance to potato late blight, Proc. Natl. Acad. Sci. 100 (2003) 9128–9133.

[4] S.J. Foster, T.-H. Park, M. Pel, G. Brigneti, J. Śliwka, L. Jagger, E. van der Vossen, J.D.G. Jones, Rpi-vnt1.1, a Tm-22 homolog from Solanum venturii, confers resistance to Potato Late Blight, Mol. Plant Microbe Interact. 22 (2009) 589–600, https://doi.org/10.1094/MPMI-22-5-0589.

[5] E. Runno-Paurson, I.H. Williams, L. Metspalu, T. Kaart, M. Mänd, Current potato varieties are too susceptible to late blight to be grown without chemical control under North-East European conditions, Acta Agric. Scand. Sect. B — Soil Plant Sci. 63 (2013) 80–88, https://doi.org/10.1080/09064710.2012.721389.

[6] W.H. Bohl, S.B. Johnson, Commercial Potato Production in North America, (n.d.) 90.

[7] P.J. Curran, Remote sensing of foliar chemistry, Remote Sens. Environ. 30 (1989) 271–278, https://doi.org/10.1016/0034-4257(89)90069-2.

[8] R.F. Kokaly, G.P. Asner, S.V. Ollinger, M.E. Martin, C.A. Wessman, Characterizing canopy biochemistry from imaging spectroscopy and its application to ecosystem studies, Remote Sens. Environ. 113 (2009) S78–S91, https://doi.org/10.1016/j.rse.2008.10.018.

[9] R.F. Kokaly, A.K. Skidmore, Plant phenolics and absorption features in vegetation reflectance spectra near 1.66μm, Int. J. Appl. Earth Obs. Geoinf. 43 (2015) 55–83, https://doi.org/10.1016/j.jag.2015.01.010.

[10] S.P. Serbin, Spectroscopic Determination of Leaf Nutritional, Morpholgical, and Metabolic Traits, PhD Thesis, (2012) http://adsabs.harvard.edu/abs/2012PhDT.....189S (Accessed June 14, 2019).

[11] A. Singh, S.P. Serbin, B.E. McNeil, C.C. Kingdon, P.A. Townsend, Imaging spectroscopy algorithms for mapping canopy foliar chemical and morphological traits and their uncertainties, Ecol. Appl. 25 (2015) 2180–2197, https://doi.org/10.1890/14-2098.1.

[12] S.P. Serbin, A. Singh, B.E. McNeil, C.C. Kingdon, P.A. Townsend, Spectroscopic determination of leaf morphological and biochemical traits for northern temperate and boreal tree species, Ecol. Appl. 24 (2014) 1651–1669, https://doi.org/10.1890/13-2110.1.

[13] J.J. Couture, S.P. Serbin, P.A. Townsend, Spectroscopic sensitivity of real-time, rapidly induced phytochemical change in response to damage, New Phytol. 198 (2013) 311–319, https://doi.org/10.1111/nph.12159.

[14] J.J. Couture, A. Singh, K.F. Rubert-Nason, S.P. Serbin, R.L. Lindroth, P.A. Townsend, Spectroscopic determination of ecologically relevant plant secondary metabolites, Methods Ecol. Evol. 7 (2016) 1402–1412, https://doi.org/10.1111/2041-210X.12596.

[15] G.P. Asner, R.E. Martin, Spectral and chemical analysis of tropical forests: scaling from leaf to canopy levels, Remote Sens. Environ. 112 (2008) 3958–3970.

[16] J. Cavender-Bares, J.E. Meireles, J.J. Couture, M.A. Kaproth, C.C. Kingdon, A. Singh, S.P. Serbin, A. Center, E. Zuniga, G. Pilz, P.A. Townsend, Associations of leaf spectra with genetic and phylogenetic variation in oaks: prospects for remote detection of biodiversity, Remote Sens. 8 (2016) 221, https://doi.org/10.3390/rs8030221.

[17] M. Yuan, J.J. Couture, P.A. Townsend, M.D. Ruark, W.L. Bland, Spectroscopic determination of leaf nitrogen concentration and mass per area in sweet corn and snap bean, Agron. J. 108 (2016) 2519–2526, https://doi.org/10.2134/agronj2016.05.0260.

[18] Z. Seregély, T. Deák, G.D. Bisztray, Distinguishing melon genotypes using NIR spectroscopy, Chemometr. Intell. Lab. Syst. 72 (2004) 195–203, https://doi.org/10.1016/j.chemolab.2004.01.013.

[19] I. Esteban-Díez, J.M. González-Sáiz, C. Pizarro, Prediction of sensory properties of espresso from roasted coffee samples by near-infrared spectroscopy, Anal. Chim. Acta 525 (2004) 171–182, https://doi.org/10.1016/j.aca.2004.08.057.

[20] M.A. Brescia, G.D. Martino, C. Fares, N.D. Fonzo, C. Platani, S. Ghelli, F. Reniero, A. Sacco, Characterization of italian durum wheat semolina by means of chemical analytical and spectroscopic determinations, Cereal Chem. 79 (2002) 238–242, https://doi.org/10.1094/CCHEM.2002.79.2.238.

[21] T. Behrens, M. Kraft, F. Wiesler, Influence of measuring angle, nitrogen fertilization, and variety on spectral reflectance of winter oilseed rape canopies, J. Plant Nutr. Soil Sci. 167 (2004) 99–105.

[22] C.S.T. Daughtry, C.L. Walthall, Spectral discrimination of Cannabis sativa L. leaves and canopies, Remote Sens. Environ. 64 (1998) 192–201, https://doi.org/10.1016/S0034-4257(98)00002-9.

[23] X. Li, Y. He, Discriminating varieties of tea plant based on Vis/NIR spectral characteristics and using artificial neural networks, Biosyst. Eng. 99 (2008) 313–321, https://doi.org/10.1016/j.biosystemseng.2007.11.007.

[24] N.R. Rao, P.K. Garg, S.K. Ghosh, Development of an agricultural crops spectral library and classification of crops at cultivar level using hyperspectral data, Precis. Agric. 8 (2007) 173–185, https://doi.org/10.1007/s11119-007-9037-x.

[25] G. Agrios, Plant Pathogens and Disease: General Introduction, (2009).

[26] S. Delalieux, J. van Aardt, W. Keulemans, E. Schrevens, P. Coppin, Detection of biotic stress (Venturia inaequalis) in apple trees using hyperspectral data: non-parametric statistical approaches and physiological implications, Eur. J. Agron. 27 (2007) 130–143, https://doi.org/10.1016/j.eja.2007.02.005.

[27] X. Cao, Y. Luo, Y. Zhou, X. Duan, D. Cheng, Detection of powdery mildew in two winter wheat cultivars using canopy hyperspectral reflectance, Crop Prot. 45 (2013) 124–131, https://doi.org/10.1016/j.cropro.2012.12.002.

[28] E.-C. Oerke, K. Herzog, R. Toepfer, Hyperspectral phenotyping of the reaction of grapevine genotypes to Plasmopara viticola, J. Exp. Bot. 67 (2016) 5529–5543, https://doi.org/10.1093/jxb/erw318.

[29] P. Baranowski, W. Mazurek, J. Wozniak, U. Majewska, Detection of early bruises in apples using hyperspectral data and thermal imaging, J. Food Eng. 110 (2012) 345–355.

[30] A. Siedliska, P. Baranowski, W. Mazurek, Classification models of bruise and cultivar detection on the basis of hyperspectral imaging data, Comput. Electron. Agric. 106 (2014) 66–74, https://doi.org/10.1016/j.compag.2014.05.012.

[31] J. Stark, S. Love, Potato production systems, Mosc. ID Univ. Ida. Agric. Commun. (2003).

[32] J.J. Couture, A. Singh, A.O. Charkowski, R.L. Groves, S.M. Gray, P.C. Bethke, P.A. Townsend, Integrating spectroscopy with potato disease management, Plant Dis. 102 (2018) 2233–2240, https://doi.org/10.1094/PDIS-01-18-0054-RE.

[33] Y. Zhai, L. Cui, X. Zhou, Y. Gao, T. Fei, W. Gao, Estimation of nitrogen, phosphorus, and potassium contents in the leaves of different plants using laboratory-based visible and near-infrared reflectance spectroscopy: comparison of partial least-square regression and support vector machine regression methods, Int. J. Remote Sens. 34 (2013) 2502–2518, https://doi.org/10.1080/01431161.2012.746484.

[34] J. Oksanen, F.G. Blanchet, R. Kindt, P. Legendre, P.R. Minchin, R. O'hara, G.L. Simpson, P. Solymos, M.H.H. Stevens, H. Wagner, et al., Package 'vegan,' Community Ecol. Package Version. 2, (2013).

[35] M.J. Anderson, D.C.I. Walsh, Permanova, Anosim, and the mantel test in the face of heterogeneous dispersions: what null hypothesis are you testing? Ecol. Monogr. 83 (2013) 557–574, https://doi.org/10.1890/12-2010.1.

[36] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32, https://doi.org/10.1023/A:1010933404324.

[37] B.-H. Mevik, R. Wehrens, K.H. Liland, pls: Partial Least Squares and Principal Component Regression, R Package Version. 2, (2011).

[38] M. Kuhn, Caret: classification and regression training, Astrophys. Source Code Libr. (2015) ascl:1505.003.

[39] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, Chemometr. Intell. Lab. Syst. 58 (2001) 109–130, https://doi.org/10.1016/S0169-7439(01)00155-1.

[40] B. Worley, R. Powers, Multivariate analysis in metabolomics, Curr. Metabolomics 1 (2013) 92–107.

[41] I. Herrmann, S.K. Vosberg, P. Ravindran, A. Singh, H.-X. Chang, M.I. Chilvers, S.P. Conley, P.A. Townsend, Leaf and canopy level detection of fusarium virguliforme (sudden death syndrome) in soybean, Remote Sens. 10 (2018) 426, https://doi.org/10.3390/rs10030426.

[42] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data

Mining, Inference, and Prediction, Springer Series in Statistics, Springer, New York, 2009.
[43] M. Immitzer, C. Atzberger, T. Koukal, Tree species classification with random forest using very high spatial resolution 8-band worldview-2 satellite data, Remote Sens. 4 (2012) 2661–2693, https://doi.org/10.3390/rs4092661.
[44] A. Stevens, L. Ramirez–Lopez, An Introduction to the Prospectr Package, R Package Vignette Rep. No R Package Version 01. 3, (2014).
[45] T.H. Demetriades-Shah, M.D. Steven, J.A. Clark, High resolution derivative spectra in remote sensing, Remote Sens. Environ. 33 (1990) 55–64, https://doi.org/10.1016/0034-4257(90)90055-Q.
[46] J. Cohen, A coefficient of agreement for nominal scales, Educ. Psychol. Meas. 20 (1960) 37–46, https://doi.org/10.1177/001316446002000104.
[47] H. Poorter, Ü. Niinemets, L. Poorter, I.J. Wright, R. Villar, Causes and consequences of variation in leaf mass per area (LMA): a meta-analysis, New Phytol. 182 (2009) 565–588, https://doi.org/10.1111/j.1469-8137.2009.02830.x.
[48] B. Gao, NDWI—a normalized difference water index for remote sensing of vegetation liquid water from space, Remote Sens. Environ. 58 (1996) 257–266, https://doi.org/10.1016/S0034-4257(96)00067-3.
[49] J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, R.D.C. Team, et al., nlme: linear and nonlinear mixed effects models. R Package Version 3.1–111, Comput. Man. (2013).
[50] R. Lenth, Emmeans: Estimated Marginal Means, Aka Least-squares Means, R Package Version. 1, (2018).
[51] K.M. Gold, A.J. Gevens, P.A. Townsend, System For Detection Of Disease In Plants, US Patent App, 2019 16/251,415.
[52] M.A. Henriquez, L.R. Adam, F. Daayf, Alteration of secondary metabolites' profiles in potato leaves in response to weakly and highly aggressive isolates of Phytophthora infestans, Plant Physiol. Biochem. 57 (2012) 8–14, https://doi.org/10.1016/j.plaphy.2012.04.013.
[53] D.A. Halterman, Y. Chen, J. Sopee, J. Berduo-Sandoval, A. Sánchez-Pérez, Competition between Phytophthora infestans effectors leads to increased aggressiveness on plants containing broad-spectrum late blight resistance, PLoS One 5 (2010) e10536, , https://doi.org/10.1371/journal.pone.0010536.
[54] S. Thomas, J. Behmann, A. Steier, T. Kraska, O. Muller, U. Rascher, A.-K. Mahlein, Quantitative assessment of disease severity and rating of barley cultivars based on hyperspectral imaging in a non-invasive, automated phenotyping platform, Plant Methods 14 (2018) 45, https://doi.org/10.1186/s13007-018-0313-8.
[55] K. Gold, Contact Reflectance Spectroscopy for Accurate Phytophthora infestans Clonal Lineage Differentiation, (2019) in press.
[56] D.C. Erwin, O.K. Ribeiro, Phytophthora Diseases worldwide., Phytophthora dis. Worldw, (1996) (Accessed June 14, 2019), https://www.cabdirect.org/cabdirect/abstract/19971001256.
[57] R.L. Nicholson, R. Hammerschmidt, Phenolic compounds and their role in disease resistance, Annu. Rev. Phytopathol. 30 (1992) 369–389, https://doi.org/10.1146/annurev.py.30.090192.002101.
[58] X. Wang, A. El Hadrami, L.R. Adam, F. Daayf, Differential Activation and Suppression of Potato Defence Responses by Phytophthora infestans Isolates Representing US-1 and US-8 Genotypes, (2008), pp. 1026–1037, https://doi.org/10.1111/j.1365-3059.2008.01866.x.
[59] P. Galeano Garcia, F. Neves dos Santos, S. Zanotta, M.N. Eberlin, C. Carazzone, Metabolomics of Solanum lycopersicum infected with Phytophthora infestans leads to early detection of late blight in asymptomatic plants, Molecules 23 (2018) 3330, https://doi.org/10.3390/molecules23123330.
[60] R.H.J. Heim, I.J. Wright, H.-C. Chang, A.J. Carnegie, G.S. Pegg, E.K. Lancaster, D.S. Falster, J. Oldeland, Detecting Myrtle Rust (Austropuccinia Psidii) on Lemon Myrtle Trees Using Spectral Signatures and Machine Learning, Plant Pathol. 67 (5) (2018) 1114–1121, https://doi.org/10.1111/ppa.12830.