Use of Near-Infrared Reflectance Spectrometry and Multivariate Data Analysis to Detect Anther Smut Disease (*Microbotryum violaceum*) in *Silene dioica*

M. Nilsson, T. Elmqvist, and U. Carlsson

First author: Department of Forest Ecology, Swedish University of Agricultural Sciences, S-901 83 Umeå; second and third authors: Department of Ecological Botany, University of Umeå, S-901 87 Umeå, Sweden.

We thank P. Geladi for valuable comments on the manuscript, C. Jonsson for technical assistance in the lab, and two anonymous referees who significantly improved the manuscript.

The project received financial support from the Swedish Natural Research Council.

Accepted for publication 9 March 1994.

ABSTRACT

Nilsson, M., Elmqvist, T., and Carlsson, U. 1994. Use of near-infrared reflectance spectrometry and multivariate data analysis to detect anther smut disease (*Microbotryum violaceum*) in *Silene dioica*. Phytopathology 84:764-770.

Near-infrared reflectance (NIR) spectral data was used in principal component analysis (PCA) to detect infection of *Silene dioica* by *Microbotryum violaceum*. Rosette leaf samples were accurately identified as either healthy (97%) or infected (96%) when NIR data was analyzed by PCA. The two classes overlapped slightly when principal component models were used to classify unknown samples. A method to measure the degree of infection is also presented. The use of NIR and PCA for both detection and quantification of fungal biomass in plant material should be useful for studying plant-pathogen interactions and as a method for assessing disease incidence in crops.

Additional keywords: chitin, pattern recognition.

In plant populations, it is often very difficult to determine the incidence of infection caused by fungal pathogens. Pathogens often cause apparent symptoms of infection only after a period of incubation (9,24,36). Symptoms may be present only in specific tissues, and some infections may be completely asymptotic (6,8). In order to model plant-pathogen dynamics, it is essential to know the actual incidence of plant infection in a population (i.e., the proportion of plants infected) in contrast to disease incidence, as determined by visual inspection.

Different strategies for detection of plant pathogens have been utilized. The oldest and most common method is histological examination of stained preparations (20,26,29). This method is quite reliable in some cases but not in others. Chemical quantification of compounds specific to fungal pathogens, such as chitin (27,28,35) and ergosterol (14,23,33), has been used with good results. Whipps et al (37) provided an excellent review of the use of cell wall compounds, storage carbohydrates, and enzymes as biochemical markers for fungal pathogens. They pointed out that for quantitative analyses, a thorough study of the variation in fungal components during the different phases of the host-parasite interaction is needed. DNA hybridization (13) and immunofluorescence (17) are techniques that have been used to assess the presence of fungi in plant material. Both techniques give good results, but they are highly species specific and therefore need development for every new pathogen studied.

Only a few attempts have been made to use absorbance in the near-infrared region as a means to detect the presence of plant pathogens. Birth (5) constructed a "smut meter" that determined the number of spores in wheat. Another approach was used by Asher et al (2), who collected spores from fungal pathogens on glass-fiber disks. Absorbances at three different wavelengths were measured and highly correlated (>0.9) with actual spore counts made with a hemacytometer. For potato tubers, reflectance in the near-infrared region was found to be altered by certain diseases and defects (25). Detection and quantification of molds in hay and cereals have been performed with near-infrared reflectance (NIR) spectroscopy calibrated by high-performance liquid chromatography to determine chitin content and different amounts of fungal mycelium mixed with barley grains (30–32).

Organic chemical compounds can be analyzed by measuring the absorbance spectrum in the near-infrared range with NIR spectrometry (21,22). The main constituents of organic material (water, fat, protein, and carbohydrates) show distinct, although overlapping, radiation absorption spectra in the NIR region (700–2,500 nm) (22). The resulting spectra of complex organic materials are composed of overlapping spectra specific for every chemical constituent and in most cases do not reveal any visually

interpretable information by themselves. To gain more interpretable information from NIR data, multivariate data analysis can be used (16,18), either for discrimination between samples or for prediction of certain constituents.

Silene dioica (L.) Clairv. (Caryophyllaceae) is a dioecious perennial herb that is widespread in northern Fennoscandia (Norway, Sweden, and Finland). It occurs in wet, mesic, moderately to richly fertile soils found in meadows, deciduous and coniferous forests, and the deciduous phase of primary successions along the shore of the Gulf of Bothnia. It may produce several erect flowering shoots, and the basal rosette leaves persist through the winter. New leafy shoots are produced during the summer as leafy stolons. Flowering begins in late May or early June (males a few days before females), and seeds are dispersed in mid- to late July.

Populations of *S. dioica* are frequently infected by *Microbotryum violaceum* (Brandenburger and Schwinn) G. Deml. and Oberwinkler (Ustilaginales) (syn. *Ustilago violacea*). The disease is systemic and results in sterility in both male and female plants. Teliospores are transmitted mainly by pollinating insects (4). Spores deposited on healthy flowers germinate and undergo meiosis to produce haploid sporidia of opposite mating types. These opposite mating types conjugate, resulting in dikaryotic hyphae that invade the host. The fungus grows through the plant and survives the winter in the root. During the next season, all flowers on the infected plant exhibit disease symptoms. Infection results in a striking change in floral morphology, and stamens with anther sacs filled with fungus spores develop in both male and female flowers (4).

The objectives of this study were 1) to develop a method that used both NIR spectrometry and principal component analysis (PCA) to detect the presence of fungal plant pathogens in plant tissues and 2) more specifically, to develop a calibration model based on NIR absorbance spectra, which discriminate between healthy and infected plants of *S. dioica.* In *S. dioica,* infection is visually detectable only in plants in the flowering stage. With the existing methods, mycelium can be found abundantly in meristematic tissues close to flowers and sparingly in shoots or leaves away from the flowers (15). Audran and Batcho (3) concluded that when only vegetative tissues are available, it is difficult with existing techniques to separate infected from noninfected *S. dioica.*

MATERIALS AND METHODS

Plants. The plants used in this study were collected from three different field populations located on the coast of the Gulf of Bothnia (64°14′N, 22°16′E) during the summer of 1988. They were classified as healthy or infected according to the presence or absence of teliospores in flowers. Only flowering plants with either all flowers infected or all flowers healthy were collected.

NIR spectroscopy. A model 400 D/R InfraAlyzer (Technicon Industrial Systems, New York, NY), operating with 19 different optical filters from 1,445 to 2,348 nm, was used to collect NIR spectra of the samples. The following wavelengths (nanometers) were used: 1,445, 1,680, 1,722, 1,734, 1,759, 1,778, 1,818, 1,940, 1,982, 2,100, 2,139, 2,180, 2,190, 2,208, 2,230, 2,270, 2,310, 2,336, and 2,348. Plant samples were dried for 48 h at 40 C and milled in a ball mill for 1 min. Samples of approximately 1 ml were used to obtain spectral data.

Plant material for data sets I and II were collected in late July (late flowering) and material for set III in early June (early flowering) (Fig. 1). Data set IV was derived from plants in data set II. It comprises 11 mixtures with increasing amounts of infected leaves as follows: sample 1, 0%; 2, 9.6%; 3, 21.4%; 4, 28.3%; 5, 39.4%; 6, 49.4%; 7, 60.3%; 8, 70.0%; 9, 80.1%; 10, 91.0%, and 11, 100%.

Data analysis. Samples were classified into one of four groups according to tissue type (stem or leaves) and infection (present or absent) (Fig. 1). The statistical analysis was executed in three phases: 1) sorting the samples into classes according to similarities in measured NIR variables; 2) describing the classes by principal component models; and 3) validating the class models with the use of a set of reference samples. Outliers (i.e., single samples that deviated from any of the sample populations) can be detected by analysis of leverage, by unmodeled variance, and/or by plotting the component scores. They were recognized in the principal component score plots and removed after analysis phase 1. The effect of outliers in either calibration or test data sets is discussed by Martens and Ness (19). The effect of different light scattering from the samples was corrected by using mean value scatter correction (11). For multivariate data analysis, the SIMCA software (Umetri AB, Umeå, Sweden) was used (39–41) on an IBM-AT personal computer. A general description of PCA can be found elsewhere (16,18). Below, only the special multivariate procedures used in the actual data analysis are outlined.

The data (absorbances at each wavelength) for each class (one of the data sets I–IV in Figure 1) were approximated by a separate principal component model. The class limits were determined by the standard deviation (SD) of the residuals between the fitted variables of the samples and the measured ones. New, unclassified NIR absorbance spectra were then classified according to their degree of fit to the different class models (i.e., whether or not the distance between the sample and the model fell inside the 95% confidence interval of the class limits) by multiple linear regression (1,40). The data were analyzed sequentially, as outlined in Figure 1 and Table 1. Each class was scaled separately (1) and described by the principal components of its NIR spectra. The number of significant (CSV/SD < 0.95, where CSV is the cross-validated SD after the actual component and SD is the residual standard deviation before computation of the actual component) components in each class model was determined by

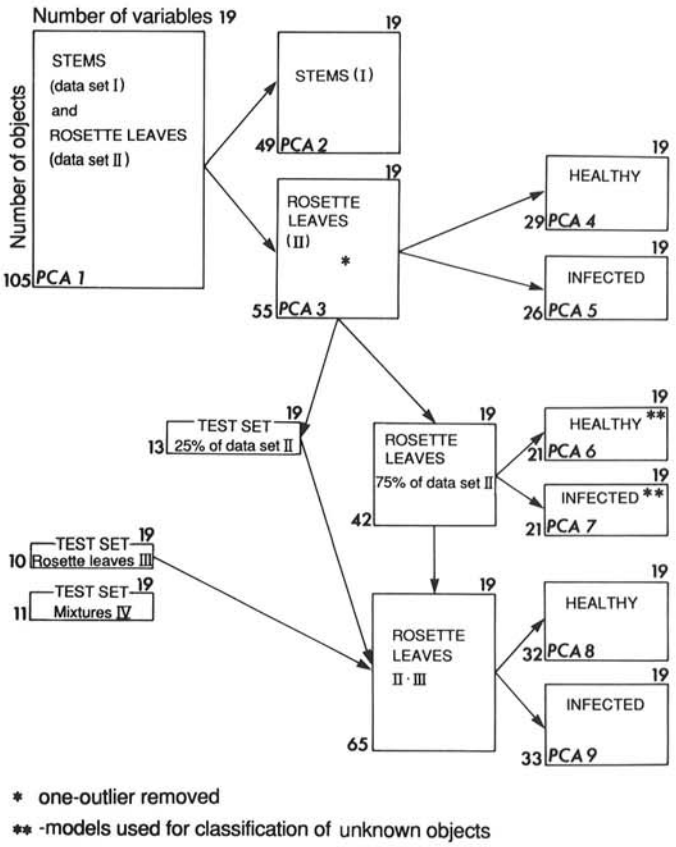

Fig. 1. Flow chart of the data sets, steps in the principal component analysis (PCA), and subsequent classification of unknown samples. The arrows show how the data sets were partitioned for further analysis. Data sets I and II are from the same sampling occasion. Data set III is from the same population but was collected 3 mo earlier, and set IV consists of 11 mixtures of rosette leaves from one infected and two healthy plants. The number at the upper right corner of each box is the number of variables (wavelengths), and the number at the lower left is the number of samples. PCA numbers refer to the model numbers listed in Table 1.

cross-validation (38). If all significant components were used, the model of the reference set would be overfitted (20) and samples in the test set would be classified outside the 95% confidence limits of its own class. The optimal number of components used for classification of unknown samples was determined by comparison of the class SD of the reference set with the class SD of the test set. Additional models for both the reference and the test sets were calculated by increasing the number of significant principal components. The number of components selected was based on the SD of the test set that most closely approximated that of the reference set.

The class limits used for classification of new samples were calculated as follows. One fourth of the samples were excluded at a time, and a new model was calculated from the remaining samples. The excluded samples were reclassified by the model. This procedure was repeated until all samples in the calibration set were reclassified. From the SD of the reclassified sample, a mean SD was calculated. This SD was obtained from the residuals between classified and measured variables and therefore was more related to the SD of new, unknown samples (34). *F* values were used to test whether the classified samples significantly deviated from the model; confidence intervals of 95% were used.

RESULTS

Principal component models accounted for more than 95% of the variance in absorbance in diseased and healthy plants of *S. dioica* (Table 1). When all samples were analyzed by principal component analysis model 1 (PCA 1; PCAs 1–9 refer to models in Table 1), the data were clearly separated into stems and rosette leaves (Fig. 2A). A less clear separation was seen between healthy and infected samples. In Figure 2A, one sample from the class of healthy leaves was revealed as an outlier and removed in the subsequent data analyses. According to the separation obtained in PCA 1, the data from stems (set I; data sets I–IV refer to Figure 1 and Table 1) and rosette leaves (set II) were further treated separately. In the PCA of stems only (PCA 2), the two first components explained 66% of the variance; however, both components only partly distinguished infected and noninfected stems (Fig. 2B). In the analysis of data from the rosette leaves (set II, PCA 3), infected leaves were distinguished from healthy leaves (Fig. 2C and D). Thus, data from the rosette leaves (set II) were chosen for further development of the detection of *M. violaceum* in *S. dioica*.

In phase 2, one model was computed for the healthy class (PCA 4) and one for the infected class (PCA 5). Each class was described with seven significant components. The samples were expected to have a short distance (i.e., the SD of the residuals between the fitted and measured variables of the sample) to their own class model and a longer distance to the model of the other classes. The distances to the models for healthy and infected leaves (PCAs 4 and 5) for every sample in data set II (rosette leaves) are plotted in Figure 3A. The 95% confidence limit for the classes is also indicated. One healthy and one infected sample were classified inside the 95% confidence limit for the wrong class.

Classification of unknown samples. To test the accuracy of the classification properties of the models, three different sets of data were used (Fig. 1). Twenty-five percent of the rosette leaves (set II) were removed, and two new models (PCAs 6 and 7) were developed on the basis of the remaining 75% of the samples. The SDs for all samples included in the models are plotted in Figure 3B. The models (PCAs 6 and 7) computed on this data set were used for further classification of three sets of unknown samples, shown in Figure 3C. The removed samples (25%) were fitted to the class models of healthy (PCA 6) and infected (PCA 7) samples (Fig. 3C). Another set of samples ($n = 10$) from plants collected earlier from the same plant populations (set III) were also classified by PCAs 6 and 7 (Fig. 3C).

Of the samples that had been removed from the original data set (rosette leaves of set II), three healthy samples and one infected sample were classified to both classes, and one sample was fitted outside the class of healthy leaves. Of the samples of rosette leaves collected in early June (test set III), all but one of the healthy and two of the infected samples were classified as both healthy and infected. One healthy sample was outside its own class.

The two test sets used as unknown samples (Fig. 1) were included when PCAs 8 and 9 were calculated (Fig. 3D). The plotted samples are equal to those in Figure 3C, but in this figure they are classified by PCAs 6 and 7. The two plots are similar, but slightly better discrimination was obtained when all samples were included for model development. The relative distances (i.e., the residual variance for the sample not described by the model) between the

TABLE 1. Principal component models developed from different data sets of stems and rosette leaves of *Silene diocea*

Principal component analysis model	Data set number	Principal component[a]							Sum of explained variance (component used)
		C1	C2	C3	C4	C5	C6	C7	
1	Stems and leaves, I and II	0.70	0.60	0.80	0.73	0.64	0.89	ns	
		49.6	33.6	7.5	4.8	2.9	0.6	...	99.0 (6)
2	Stems, I	0.73	0.88	0.66	0.72	0.74	0.93	ns	
		47.0	19.3	19.7	8.0	3.0	0.9	...	97.9 (6)
3	Rosette leaves, II	0.68	0.61	0.73	0.67	0.90	ns	ns	
		55.9	27.9	8.1	5.0	1.0	...	...	97.9 (5)
4	Rosette leaves, II healthy	0.79	0.72	0.85	0.70	0.91	0.82	0.83	
		47.2	29.5	8.5	8.9	1.9	1.3	1.0	98.5 (7)
5	Rosette leaves, II infected	0.82	0.74	0.73	0.75	0.85	0.85	0.85	
		42.1	29.7	14.1	7.9	2.6	1.6	0.9	98.9 (7)
6	75% of rosette leaves, II infected	0.71	0.71	0.98	0.83	0.85	ns	ns	
		51.4	26.9	9.1	6.6	2.6	...	...	97.7 (5)
7	75% of rosette leaves, II infected	0.82	0.71	0.82	0.78	0.80	0.86	ns	
		43.7	28.7	11.1	8.9	3.5	2.0	...	97.9 (6)
8	Rosette leaves, II and III healthy	0.75	0.64	0.78	0.76	0.91	0.87	0.88	
		46.6	33.8	9.1	5.3	1.5	1.0	1.0	98.3 (7)
9	Rosette leaves, II and III infected	0.76	0.71	0.81	0.73	0.79	ns	ns	
		44.4	28.5	11.3	8.5	3.0	...	...	95.7 (5)

[a] The first row of numbers for each data set contains the cross-validation terms for determination of component significance (CSV/SD < 0.95, where CSV = cross-validated SD after the actual component and SD = the standard deviation before calculation of the actual component); the second row contains the explained variance (%) by each component. ns = Not significant at the 95% level, and ... = no significantly explained variance.

model and the samples from infected plants were nearly the same, regardless of whether the samples were modeled or classified. The distance to the healthy model was very different and depended on whether the samples had been included in the model computation.

Quantitative response. Data set IV consisted of samples with 11 different concentrations of healthy and infected rosette leaves. These samples were classified by PCAs 6 and 7, and a plot of minimum class distance is shown in Figure 4. One severely infected sample was used for the mixtures. This sample (11 in Figure 4) is classified as infected and has by far the longest distance to the healthy model. The healthy sample used for mixtures was made of two different healthy samples of rosette leaves. The series of mixtures was positioned from the 100% infected sample along the axis for the infected model towards the model of the healthy samples. All samples were close to the model for infected leaves. The three samples with the lowest concentrations of infected leaves were classified as both infected and healthy.

Absorbance at different wavelengths. Absorbances at each wavelength were related differently to infected and healthy plants (an example from PCA 3 is shown in Table 2). The explained variance for each variable together with the chemical assignments for each wavelength are reported (22). The plus or minus sign in front of the explained variance refers to whether the variable is positively or negatively correlated to the component. The first two wavelengths with negative signs for the explained variance were the variables responsible for the negative values of component 1 in PCA 3 of all rosette leaves and were strongly correlated to infected plants. Positive values for explained variance in this first component were related to the healthy plants, i.e., positive principal component values in Figure 2C.

DISCUSSION

According to the principal component models, the NIR spectra contain information about both differences in plant tissue and the presence of *M. violaceum*. Discrimination between healthy and infected samples is most pronounced in the rosette leaves. For *S. dioica*, the NIR method successfully discriminated between healthy and diseased tissue where other methods had failed (3). In previous work, the fungal pathogen *M. violaceum* could be detected only in the stems close to the flowers (15). In this pathosystem, the NIR method will be of value in epidemiological studies. Previous ecological models have been based only on the recorded incidence among the flowering portion of a population at a particular season, although nonflowering plants may also be infected.

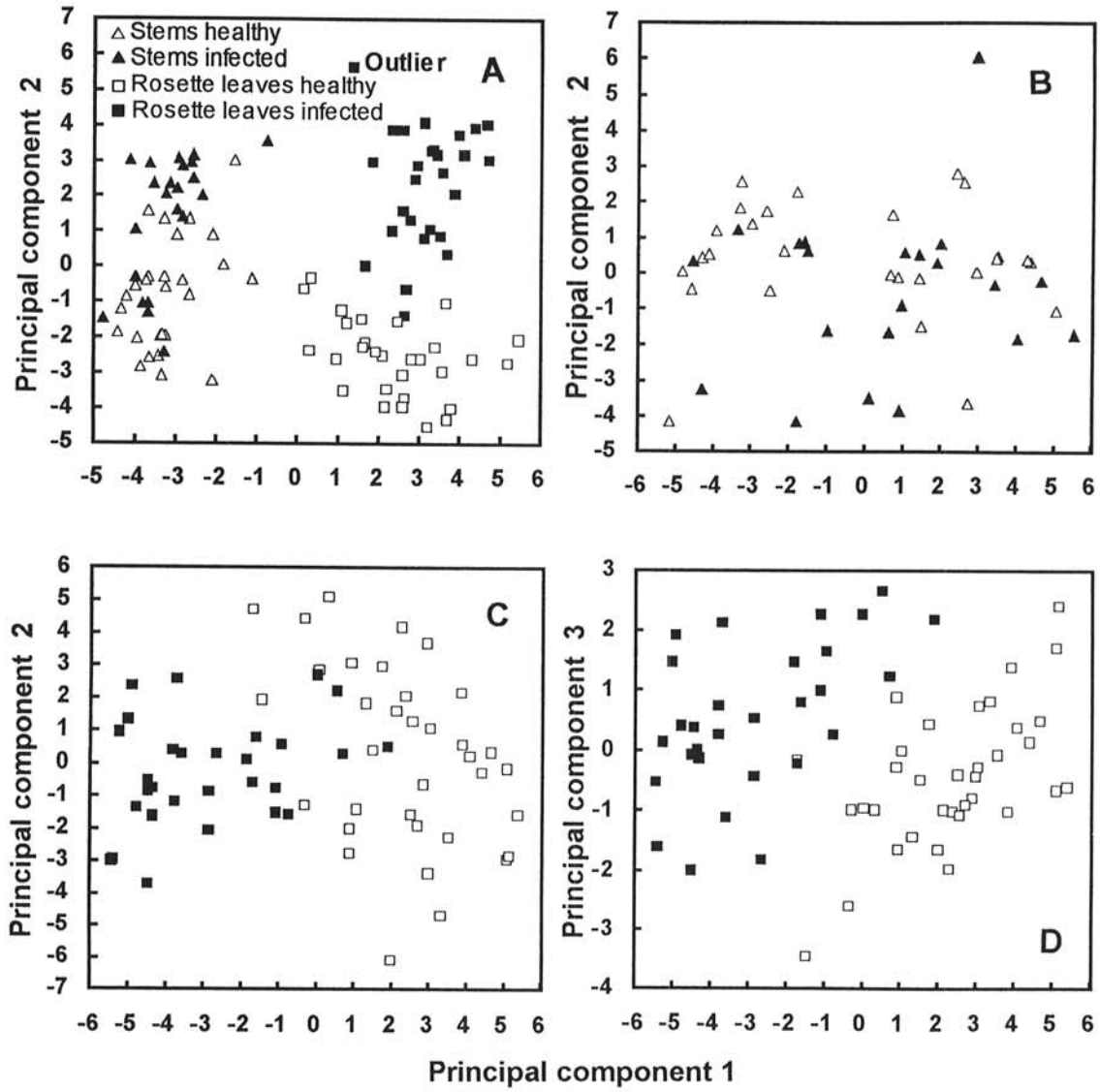

Fig. 2. Principal component scores after different data sets. Each point is one sample as determined by 19 absorbance values. **A,** Principal components 1 and 2 after principal component analysis 1 of near-infrared reflectance data sets I and II (Fig. 1) from stems and rosette leaves. **B,** Principal components 1 and 2 from healthy and infected stems (data set I). The same data on stems as in **A** were used. **C,** Principal components 1 and 2 from healthy and infected rosette leaves (data set II), with one outlier removed as indicated in **A**. **D,** Principal components 1 and 3 from healthy and infected rosette leaves (data set II).

Modeling and classification. Healthy and infected rosette leaves were accurately separated into two classes (PCAs 6 and 7, Table 1). Two infected samples were classified just inside the 95% confidence limit of the healthy samples. Classifications of the healthy samples are less accurate. Two healthy plants were excluded from their own class. Half of the samples of healthy plants were classified both to their own class and to the class of infected plants. If the samples in this test set were included in the reference set (Fig. 3D), only two infected and two healthy samples would have been placed in both classes. All others were accurately classified. This may indicate that there is variation in the plant material not covered by the reference set. If this is the case, then the problem probably can be solved by including more plants in the reference set, thereby covering more of the class-specific variation and giving a more robust model. Once a principal component model of NIR data from a certain plant-pathogen system has been established, this model can be used for classification of new, unknown samples. Eventual outliers among the new samples can be detected by any of the procedures discussed in Material and Methods.

Classification of mixtures. When the samples of mixtures in set IV were classified by PCAs 6 and 7, it was obvious that the distance between the model and the sample represented a quantitative measurement of the degree of fungus-infected material in the mixture. All samples, except the healthy and the one with the lowest proportion of infected material (10%), were classified relative to each other according to the concentration of infected plant material. The uncertainty at the lowest proportion of infected material may have resulted from poor mixing or low accuracy at this concentration.

Chemical origin of differences. The chemical differences between healthy and infected plants can be caused by either the chemical composition of the fungal parasite itself or a response from the host plant. In the latter case, we know that infection of *S. dioica* results in a sex change in females and changes in flower phenology, allocation of photosynthate (4,7), and palatability to herbivores (T. Elmqvist, *personal observation*). All these changes also imply different allocation of nutrients and energy and possibly even production of secondary metabolites. Some interpretations about the origin of discriminating chemical

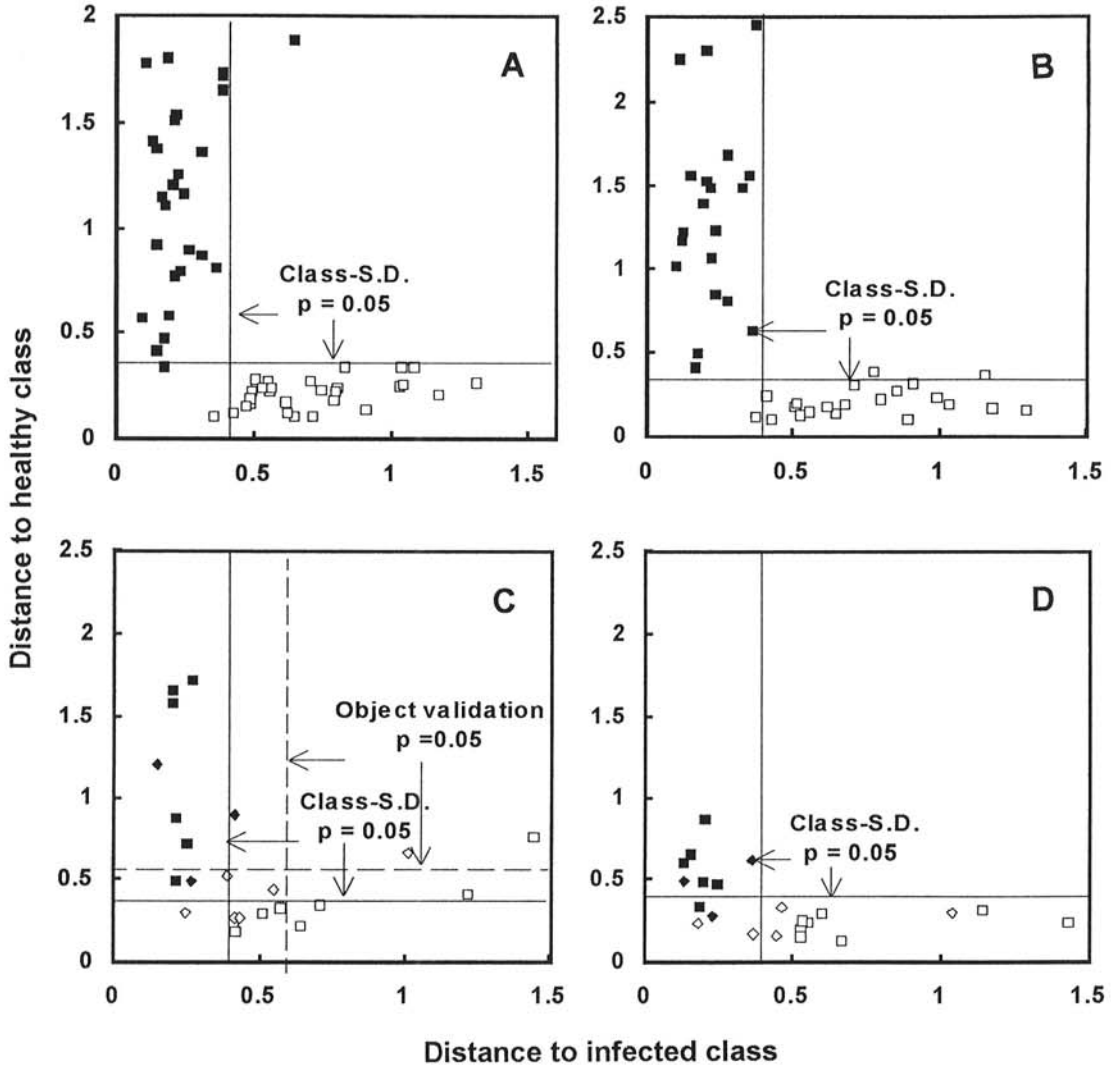

Fig. 3. The minimal class distance between each sample and class models of healthy and infected rosette leaves plotted against each other (i.e., the standard deviation [SD] for the squared residuals between the measured variables and the estimate by the model of the same variable). The distance to the model for the infected plants is plotted on the x-axis and the distance to the model for healthy plants on the y-axis. The same data set as in Figure 2C and D with rosette leaves was used. The 95% confidence limits for the models are marked with lines. **A,** Samples of rosette leaves (data set II) fitted by class models 4 and 5 are plotted against each other. Samples are the same as in Figure 2C. **B,** Samples of rosette leaves (data set II) fitted by class models 6 and 7 calculated on 75% of the data. These models were used for further classification of unknown objects. **C,** Samples of rosette leaves classified by principal component analysis (PCA) models 6 and 7. Two sets of unknown samples are classified, the 25% excluded in data set II and test set III from an earlier sampling. Dotted lines indicate 95% confidence limits of the object validation procedure. **D,** Samples are the same as in **C,** but they have been fitted by PCA models 8 and 9 together with the samples used for the classification models 6 and 7. □ = Healthy rosette leaves; ■ = infected rosette leaves from data set II; ◇ = healthy rosette leaves; and ◆ = infected rosette leaves from the earlier sampled test set III.

differences can also be made. In the analysis of data set II, most of the separation between healthy and infected plants was contained in PC 1. The influence (explained variance) from each wavelength to this separation is analyzed in Table 2. The 1,940- and 1,982-nm wavelengths appear to be important in detecting infected plants. As indicated in Table 2, amine and amide groups, as well as water, absorb energy at these wavelengths (12). Both glucose and galactose amines constitute chitin and other important cell wall components in fungi. Chitin is found in many Basidiomycetes (10), including *Microbotryum* spp. Therefore, it is likely that chitin is the compound most responsible for the NIR differences between infected and noninfected plants.

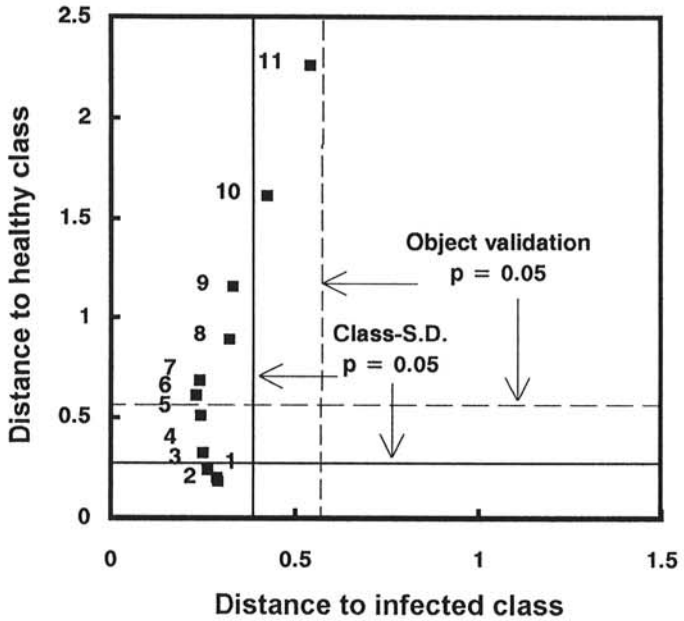

Fig. 4. Eleven mixtures of infected and healthy rosette leaves classified by principal component analysis models 6 and 7, the same models used for classification in Figure 3C. Object 1 consists of two healthy samples and 11 of a highly infected sample. The degree of infection increases by approximately 10% in each object. The plot is equal to Figure 3A–D.

The possibility of using NIR for detection of fungi provides several advantages. Large quantities of samples can be analyzed, and the same instrumentation and methods can probably be used for different plant pathogens. The NIR-PCA method accurately discriminates between healthy and infected *S. dioica* in cases where other methods have failed. The application of the method to field studies will enable researchers to estimate the actual infection incidence and not be limited to the disease incidence based on the flowering portion of the plant population. Furthermore, with this method, it will be possible to follow development of the pathogen in the plants during the season. This information will be of great value for all models on host-pathogen interaction in this and similar systems. Also, this method will enable us to determine whether there is variation in both the pathogen and the host in characters related to virulence and resistance, e.g., whether different isolates of the pathogen show differences in growth rates in hosts originating from diseased and nondiseased populations.

We suggest that the NIR-PCA method could be applicable to other pathosystems where presence or absence of a pathogen is difficult to determine macroscopically. The NIR-PCA method may prove to be superior to other methods because it is simple and fast and only small amounts of dried plant tissue are needed. The dried samples could also be used after the NIR analysis for further specific chemical analyses. Although the instruments are expensive, they recently have become very common in food and feed research. Therefore, these instruments may be readily available for many researchers in phytopathology.

LITERATURE CITED

1. Albano, C., Blomquist, G., Dunn, W. J., III, Edlund, U., Eliasson, B., Johansson, B., Nordén, B., Sjöström, M., Söderström, B., and Wold, S. 1980. Characterization and classification based on multivariate data analysis. Pages 377-386 in: Proc. Int. Congr. Pure Appl. Chem., 27th. A. Varmavuori, ed. Pergamon Press, Oxford.
2. Asher, M. J. C., Cowe, I. A., Thomas, C. E., and Cuthbertson, D. C. 1982. A rapid method of counting spores of fungal pathogens by infrared reflectance analysis. Plant Pathol. 31:363-371.
3. Audran, J. C., and Batcho, M. 1980. Localisation d'*Ustilago violacea* (Pers.) Rouss. dans le tissus de *Silene dioica* (L.) Clairv. au stade v g tatif. Ann. Phytopathol. 12:45-55.
4. Baker, H. G. 1947. Infection of species of *Melandrium* by *Ustilago violacea* (Pers.) Fuckel and the transmission of the resultant disease.

TABLE 2. Explained variance for the absorbance at each wavelength in all components from principal component analysis model 3

Wavelength[a] (nm)	Response[b]	Principal component[c,d]				
		C1 (56%)	C2 (28%)	C3 (8%)	C4 (5%)	C5 (1%)
1,940	Water, urea	−90	−0	+8	+1	−0
1,982	Amide, proteins	−80	−11	+8	+0	−0
1,445	Water	−7	+71	−10	−9	+2
2,100	Starch	+14	+57	−27	+2	−0
2,208	Urea, lactose	+16	−80	−1	−0	+1
2,190	Protein, starch	+19	−76	−4	−0	+0
2,180	Protein	+22	−68	−7	−1	+0
2,139	Protein, starch	+27	+36	−34	+2	+0
2,230	Amino acid	+32	−63	−0	+0	+1
2,348	Cellulose	+44	+28	+19	−8	+0
2,336	Fiber, ash	+56	+29	+11	−3	+1
2,310	Fat	+59	+5	+20	−11	+0
1,818	Cellulose	+62	−0	+14	+20	+5
1,778	Starch, cellulose, fiber	+65	+11	+2	+21	+0
1,680	Aromatic compounds	+81	−5	−2	−6	+0
1,759	Fat	+86	+4	+1	+7	−2
2,270	Starch, cellulose, lignin	+91	+0	−2	−1	+0
1,734	Protein	+92	−2	+0	−0	−5
1,722	Starch, cellulose	+93	−2	+0	−1	−5

[a] Wavelengths are sorted according to the explained variance in component 1.
[b] Chemical component(s) absorbing at that particular wavelength (22).
[c] Variance explained by each component is given in parentheses.
[d] Explained variance (%) for the variable by the particular component. The + or − refers to whether the variable is positively or negatively correlated to the component; i.e., samples with high negative values for component 1 have high absorbances at 1,940 and 1,982 nm, indicating high chitin content.

Ann. Bot. (London) 11:333-348.
5. Birth, G. S. 1960. Measuring the smut content of wheat. Trans. ASAE 3:19-21.
6. Boursnell, J. G. 1950. The symbiotic seed-borne fungus in the Cistacea. I. Distribution and function of the fungus in the seedling and in the tissues of the mature plant. Ann. Bot. (London) 14:217-243.
7. Carlsson, U., and Elmqvist, T. 1992. Epidemiology of anther smut disease (*Microbotryum violaceum*) and numeric regulation of populations of *Silene dioica*. Oecologia 90:509-517.
8. Clay, K. 1986. Fungal endophytes of grasses: A defensive mutualism between plants and fungi. Ecology 69:10-16.
9. Funk, A. 1985. Foliar Fungi of Western Trees. Canadian Forestry Service, Victoria, British Columbia.
10. Garraway, O. G., and Evans, R. C. 1984. Fungal Nutrition and Physiology. John Wiley & Sons, New York.
11. Geladi, P., MacDougall, D., and Martens, H. 1985. Linearization and scatter-correction for near-infrared reflectance spectra of meat. Appl. Spectrosc. 39:491-500.
12. Goddu, R. F., and Delker, D. A. 1960. Spectra-structure correlations for the near-infrared region. Anal. Chem. 32:140-141.
13. Goodwin, P. H., Kirkpatrick, B. C., and Duniway, J. M. 1989. Cloned DNA probes for identification of *Phytophthora parasitica*. Phytopathology 79:716-721.
14. Gordon, T. R., and Webster, R. K. 1984. Evaluation of ergosterol as an indicator of infestation of barley seed by *Drechslera graminea*. Phytopathology 74:1125-1127.
15. Hassan, A., and McDonald, J. A. 1971. *Ustilago violacea* on *Silene dioica*. Trans. Br. Mycol. Soc. 56:451-461.
16. Joliffe, I. T. 1986. Principal Component Analysis. Springer-Verlag, New York.
17. Kitagawa. T., Sakamoto, Y., Furumi, K., and Ogura, H. 1989. Novel enzyme immunoassays for specific detection of *Fusarium oxysporum* f. sp. *cucumerinum* and for general detection of various *Fusarium* species. Phytopathology 79:162-165.
18. Martens, H., and Naes, T. 1984. Multivariate calibration. I. Concepts and distinctions. Trends Anal. Chem. 3:204-210.
19. Martens, H., and Naes, T. 1989. Multivariate Calibration. John Wiley & Sons, New York.
20. Myers, D. F., and Fry, W. E. 1978. The development of *Gloeocercospora sorghi* in sorghum. Phytopathology 68:1147-1155.
21. Osborne, B. G. 1983. A Bibliography of Applications of Near Infrared Reflectance Spectroscopy to Food Analysis. Flour Milling and Baking Research Association, Chorleywood, Rickmansworth, England.
22. Osborne, B. G., and Fern, T. 1986. Near Infrared Spectroscopy in Food Analysis. Longman, Harlow, England.
23. Osswald, W. F., Höll, W., and Elstner, E. F. 1986. Ergosterol as a biochemical indicator of fungal infection in spruce and fir needles from different sources. Z. Naturforsch. C. Biosci. 41:542-546.
24. Parker, A. K., and Ried, J. 1969. The genus *Rhabdocline* Syd. Can. J. Bot. 47:1533-1545.
25. Porteous, R. L., Muir, A. Y., and Wastie, R. L. 1981. The identification of diseases and defects in potato tubers from measurements of optical spectral reflectance. J. Agric. Eng. Res. 26:151-160.
26. Rawlins, T. E. 1933. Phytopathological and Botanical Research Methods. John Wiley & Sons, New York.
27. Ride, J. P., and Drysdale, R. B. 1971. A chemical method for estimation of *Fusarium oxysporum* f. *lycopersici* in infected tomato plants. Physiol. Plant Pathol. 1:409-420.
28. Ride, J. P., and Drysdale, R. B. 1972. A rapid method for chemical estimation of filamentous fungi in plant tissue. Physiol. Plant Pathol. 2:7-15.
29. Riker, A. J., and Riker, R. S. 1936. Introduction to Research on Plant Diseases. John S. Swift, St. Louis, MO.
30. Roberts, C. A., Barton, F. E., II, and Moore, K. J. 1988. Estimation of *Acremonium-Coenophialum* mycelium in infected tall fescue. Agron. J. 80:737-740.
31. Roberts, C. A., Moore, K. J., Graffis, D. W., Kirby, H. W., and Walgenbach, R. P. 1987. Quantification of mold in hay by near IR reflectance spectroscopy. J. Dairy Sci. 70:2560-2564.
32. Roberts, C. A., Marquardt, R. R., Frohlich, A. A., McGraw, R. L., Rotter, R. G., and Henning, J. C. 1991. Chemical and spectral quantification of mold in contaminated barley. Cereal Chem. 68:272-275.
33. Seitz, L. M., Sauer, D. B., Burroughs, R., Mohr, H. E., and Hubbard, J. D. 1979. Ergosterol as a measure of fungal growth. Phytopathology 69:1202-1203.
34. Söderström, B., Wold, S., and Blomquist, G. 1982. Pyrolysis-gas chromatography combined with SIMCA pattern recognition for classification of fruit-bodies of some ectomycorrhizal *Suillus* species. J. Gen. Microbiol. 128:1773-1784.
35. Toppan, A., Esquerré-Tugayé, M. T., and Touzé, A. 1976. An improved approach for the accurate determination of fungal pathogens in diseased plants. Physiol. Plant Pathol. 9:241-251.
36. Weidemann, G. J., and Boone, D. M. 1984. Development of latent infections on cranberry leaves inoculated with *Botryosphaeria vaccinii*. Phytopathology 74:1041-1043.
37. Whipps, J. M., Haselwandter, K., McGee, E. E. M., and Lewis, D. H. 1982. Use of biochemical markers to determine growth, development and biomass of fungi in infected tissues, with particular reference to antagonistic and mutualistic biotrophs. Trans. Br. Mycol. Soc. 79:385-400.
38. Wold, S. 1978. Cross-validatory estimation of the number of components in factor and principal components models. Technometrics 20:397-406.
39. Wold, S., Albano, C., Dunn, W. J., III, Edlund, U., Esbensen, K., Geladi, P., Hellberg, S., Johansson, E., Lindberg, W., and Sjöström, M. 1984. Multivariate data analysis in chemistry. Pages 17-95 in: Chemometrics: Mathematics and Statistics in Chemistry. B. R. Kowalski, ed. Reidel, Dordrecht, Netherlands.
40. Wold, S., Esbensen, K., and Geladi, P. 1987. Principal component analysis. Chemometrics Intelligent Lab. Syst. 2:37-52.
41. Wold, S., and Sjöström, M. 1977. SIMCA: A method for analyzing chemical data in terms of similarity and analogy. Pages 243-282 in: Chemometrics: Theory and Application. B. R. Kowalski, ed. Symp. Ser. 52. American Chemical Society, Washington, DC.