

Letter to the Editor

Indicator Species Analysis: A Useful Tool for Plant Disease Studies

Paul M. Severns[†] and Emily M. SykesDepartment of Plant Pathology, University of Georgia, Athens, GA 30602
Accepted for publication 30 June 2020.

ABSTRACT

Indicator species analysis (ISA) uses indices of an organism's relative abundance and occurrence to estimate the strength of its associations with a priori groups of interest and a simple randomization test to evaluate the probability of association. Because ISA values tend to be greatest when a species is both relatively more abundant than other species in a particular group and it occurs more frequently in that same group (the expectations of a causal agent in diseased plants), ISA should be useful for identifying and narrowing the list of potential causal agents from a pool of pathogens in both emerging plant diseases and when the causal agent is unclear.

Recent ISA plant disease applications suggests it may either directly identify a single causal agent from a pool of potential pathogens or narrow the pool of pathogens as candidates for pathogenicity tests in the process of fulfilling Koch's postulates. In this letter, we explain the underpinnings of ISA, summarize the known applications to plant pathosystems, offer caveats about the analysis, and suggest scenarios where ISA may be broadly applicable for plant disease studies.

Keywords: ecology and epidemiology, population biology

Causal agents of emerging diseases can be difficult and time consuming to identify, especially when the symptoms are systemic and commonly shared among different pathogens. In such cases, field samples of diseased and nonsymptomatic plants are collected from the field, isolates are cultured and genotyped, pathogenicity trials are conducted, followed by successful inoculation and reisolation, to fulfill Koch's postulates. These types of efforts can be labor and time intensive, especially if there are more than a handful of potential causal agents that require isolation, pathogenicity testing, and reisolation. However, indicator species analysis (ISA), a statistical analysis frequently used by community ecologists, may help plant pathologists to either identify a single, potential causal agent, or substantially narrow the pool of potential causal agents from emerging plant diseases with equivocal symptoms.

Hill (1979) introduced the idea of using species as indicators of environmental conditions through an analysis called TWINSPAN (two-way indicator species analysis), which assigned species and samples simultaneously based on a complex correspondence analysis (CA) ordination algorithm. Almost 2 decades later, Dufrêne and Legendre (1997) introduced an alternative, straightforward, biologically intuitive, analytical alternative to TWINSPAN, ISA, which unlike TWINSPAN did not rely on a CA ordination algorithm and enabled the simultaneous evaluation of species between two or more a priori groups of interest (e.g., diseased, nondiseased, asymptomatic). Likely due to its relative analytical simplicity and wide range of applications (Siddig et al. 2016), ISA was quickly adopted and used broadly with the original manuscript (Dufrêne and Legendre 1997) accruing >7,800 citations since 1997 (from a Google Scholar search on 26 June 2020). However, ISA is rarely used in plant disease studies despite having biologically intuitive and straightforward underpinnings that should be useful to plant pathologists.

The scarcity of ISA application in plant disease studies is most likely due to a lack of awareness by plant pathologists. Although TWINSPAN could be used to evaluate a pool of potential pathogens for potential causal agents, it is a complicated and nuanced analysis compared with ISA. Ordination analyses, particularly principal components analysis (PCA) (Hotelling 1933) and canonical correspondence analysis (CCA) (ter Braak 1986), which may be more familiar to plant pathologists, could be used to select potential causal agents from a pool of pathogens. Compared with ISA, PCA and CCA ordinations often require that variables be transformed prior to analysis to ensure that they interact linearly with each other to avoid misleading solutions. However, ISA does not require additional data transformations and there are no composite ordination axes (an axis representing more than one variable simultaneously) to interpret. ISA is also a statistically less complicated analysis to perform than the aforementioned analyses, and it has a straightforward randomization test to determine whether the indicator species associated with a group could occur by chance. In the remainder of this letter we explain the underpinnings of ISA, offer why the analysis may be biologically complementary to plant disease studies, and suggest some applications that may be useful to plant pathologists.

INDICATOR SPECIES ANALYSIS

ISA simultaneously integrates relative species abundance and the frequency of occurrence to produce a maximum indicator value for each species. Each species is then assigned to the one group for which it has the greatest indicator value. Indicator values range from 0 to 100 (or 0 to 1 depending on the scaling approach). A 100 value represents a perfect indicator species, one that occurs exclusively in a group, is found in all samples from that group, and has high relative abundance within that group. A value of zero represents a species that has no indicator value for any group, it is usually either rare in the data set or occurs with a near uniform distribution in all or most groups.

An indicator value (IV) is calculated from the following formula:

$$IV_{kj} = 100(RA_{kj} \times RF_{kj})$$

IV_{kj} is the indicator value for each species j in each group k . The indicator value is a product of two proportions measures, the mean relative abundance of species in each group (RA_{kj}) and the relative

[†]Corresponding author: P. M. Severns; Paul.Severns@uga.edu

Funding: This work was supported with funding from the Hatch Act and P. M. Severns was supported by Advancing plant epidemiology for the growers of Georgia and beyond (project accession number 1023738) from the USDA National Institute of Food and Agriculture.

The author(s) declare no conflict of interest.

© 2020 The American Phytopathological Society

frequency of each species in each group (RF_{kj}), which is then multiplied by 100 to yield an indicator value.

McCune and Grace (2002) refer to RA_{kj} as a proportion measurement that represents the relative concentration of a species' abundance in a group. It is calculated as follows:

$$RA_{kj} = X_{kj} / \sum_{k=1}^g X_{kj}$$

where g = total number of groups and X = mean species abundance. For the measurement of occurrence frequency to a group (RF_{kj}), the data must first be transformed into a presence-absence matrix. Once that step is accomplished, RF_{kj} is calculated from the presence-absence matrix as follows:

$$RF_{kj} = \sum_{i=1}^{n_k} b_{ijk} / n_k$$

where n_k = the number of samples in group k , and b = species presence or absence. The highest indicator value for each species assigned to a single group from a collection of groups is denoted as IV_{max} . Significance testing for IV_{max} is accomplished through Monte Carlo (MC) randomizations (usually at least 1,000 randomizations) that reassign sample unit values. For each MC randomization, IV_{max} is recalculated and the $IV_{max} P$ value is obtained by the proportion of times the IV_{max} from MC randomizations equals or exceeds the IV_{max} from the nonrandomized, original data set.

ISA AND THE IDENTIFICATION OF DISEASE CAUSAL AGENTS

To receive a relatively high and statistically significant indicator value (IV_{max}), a species must be both loyal (exclusive) to a group and have a high relative abundance within the samples of that same group. If the pool of potential causal agents is compared between putatively diseased, asymptomatic, and plants without disease, ISA will assign each potential pathogen to a single group (diseased, asymptomatic, nondiseased) and determine whether the group association is likely to occur by chance. Biologically, we expect the causal agent or agents to occur more frequently and with greater relative abundance in diseased plants than nondiseased and asymptomatic plants. By equally weighting both occurrence and relative abundance to calculate the indicator value, plant pathologists could capitalize on ISA's ability to assign species to groups as a method to either identify a single highly probable causal agent from a pool of potential causal

agents or to focus pathogenicity tests on a small number of probable candidates. Because ISA evaluates the biologically intuitive patterns of causal agent occurrence and abundance, assignment to a group will increase in confidence (and likely biological relevance) with an increase in the number of samples in each group.

ISA APPLICATION IN PLANT PATHOLOGY

We are aware of only one instance where ISA was used to evaluate a pool of potential causal agents for an emerging, unknown disease. Rivedal et al. (2020) used ISA to filter the pool of isolates obtained from multiple tissues of infected and healthy squash plants in western Oregon. The emerging soilborne disease is characterized by stunting, root and crown rot, vascular discoloration, and late-season vine collapse. Over 10,000 isolates from several tissue types were identified morphologically (15 fungal species), and 1,783 isolates were identified to species via sequencing (86 fungal species). From this large pool of potential candidates, ISA returned a list of species with relatively high, but mostly not statistically significant, indicator values. On diseased plants, ISA produced the highest values for *Plectosphaerella cucumerina* ($IV_{max} = 36.6, P = 0.94$) and *Fusarium solani* ($IV_{max} = 35, P = 0.20$) based on morphological identification, and *P. cucumerina* ($IV_{max} = 14.4, P = 0.99$) based on isolate genotyping. ISA also produced relatively high indicator values on diseased plants for *F. oxysporum* ($IV_{max} = 47.3, P = 0.61$), *F. culmorum* ($IV_{max} = 12.3, P = 0.4$) and *Setophoma terrestris* ($IV_{max} = 15.6, P = 0.32$) based on morphological identifications, and *F. oxysporum* ($IV_{max} = 34.3, P = 0.16$), *F. solani* ($IV_{max} = 14.4, P = 0.40$), *F. culmorum* ($IV_{max} = 5.6, P = 0.03$), and *S. terrestris* ($IV_{max} = 5.7, P = 0.12$) from genotyping.

On the surface, ISA group assignments appeared to be of little use as only one species from genotyping had a statistically significant P value. However, those species with highest relative indicator values are known to have a wide host range with both pathogenic and nonpathogenic races (Aegerter et al. 2000; Arzanlou et al. 2013; Champaco et al. 1993; Chilosí et al. 2008; de Gruyter et al. 2010; García-Jiménez et al. 2008; Gwynne et al. 1997; Mehl and Epstein 2007; Zhang et al. 2006; Zuniga et al. 1997). Recognizing the high likelihood of pathogenic/nonpathogenic mixtures within the species with the greatest ISA values and that pathogenicity is unlikely to be predicted by either morphological or nucleotide sequences alone, Rivedal (2018) performed pathogenicity tests with isolates from five pathogens having the greatest ISA values (regardless of statistical significance). When each species was evaluated in isolation, disease symptoms were far less severe than those observed in the field. However, when the potential causal agents were evaluated in mixed pathogenicity trials, combinations

TABLE 1. Potential uses of indicator species analysis for plant pathologists

Purpose	Potential a priori groups	Variables (species)
Identify potential causal agent from a pool of pathogens for a disease with ambiguous and/or widespread symptoms	Symptomatic/nonsymptomatic plants Diseased/nondiseased plants	Number of isolates, species, or genotypes (even phenotype such as fungicide resistant) in each plant from each group
Are specific diseases associated with a specific cultivar?	Host cultivar lines	Disease relative abundance and/or occurrence on replicates of each cultivar
Are different cultivation practices (fertilizer, irrigation, crop rotations, fungicides, etc.) associated with a specific plant disease?	Fields (or experimental plots) with different treatments	Relative abundance/occurrence of all diseases
Selection of potential insect vectors	Fields, experimental plots, or plants with disease symptoms or asymptomatic	Relative abundance of all insect species trapped or actively sampled within replicates
In what host tissue types are potential causal agents most likely to occur?	Leaves, flowers, stems, roots, etc. of diseased/nondiseased plants	Relative abundance of all pathogens occurring in each tissue type of diseased and nondiseased plants
Use known indicator species to gauge changes in growing conditions due to environmental thresholds (climate, water deficit, fertilizer concentration, soil conditions, etc.)	Daily maximum temperature, growing degree day intervals, % soil moisture, soil cation exchange capacity, etc.	Relative abundance and occurrence of plant diseases (including the known indicator species) in sampling units over space and time

of causal agents, particularly combinations with *F. solani*, produced symptom ratings similar to those recorded in the field (Rivedal 2018). Based on ISA results and pathogenicity trials, the emerging vine collapse disease appeared to be caused by a fungal pathogen complex (Rivedal 2018), a known phenomenon in cucurbits (Aegerter et al. 2000; Carlucci et al. 2012; Chilosì et al. 2008; Martyn and Miller 1996).

The example of ISA application by Rivedal et al. (2020) represents a scenario in which, retrospectively, the signal from the true causal agents would obviously be confounded by a pathogen complex comprised of fungal species known to harbor both pathogenic and nonpathogenic races. Even in this data set, where the causal agent signal is obscured by biological and taxonomic ambiguity, ISA appeared useful in focusing pathogenicity testing efforts by identifying the more likely causal agents. In a more straightforward application of ISA, Severns et al. (2020) reanalyzed soil nematode samples from production blueberry fields in Georgia and North Carolina, U.S.A. (Jagdale et al. 2013) to evaluate whether there may be plant pathogenic nematode species that were missed by the traditional, abundance-based, methods of assessment. Candidate pathogenic nematodes (*Dolichodorus* spp. and *Hemicyclophora* spp.) were identified for North Carolina blueberries through a combination of nonmetric multidimensional scaling analysis (Kruskal 1964), multirank permutation procedure (Mielke and Berry 2001), and ISA (Severns et al. 2020). Importantly, ISA identified ring nematode (*Mesocriconema odoratum*) as a strong, statistically significant indicator of Georgia blueberries in the soil samples ($IV_{max} = 80.0$, $P = 0.0001$) from a pool of co-occurring nematodes (*Belonolaimus* spp., *Helicotylenchus* spp., *Hemicyclophora* spp., *Hoplolaimus* spp., *Paratrichodoros* spp., *Tylenchus* spp., and *Xiphinema* spp.). This result corroborated container studies confirming *M. odoratum* pathogenicity to Georgia highbush blueberry plants (Jagdale et al. 2013).

ADDITIONAL ISA APPLICATIONS AND EXTENSIONS

The components of ISA, estimates of species concentration (RA_{kj}) and occurrence (RF_{kj}), provide information that may indicate the nature of pathogen and causal agent associations with different groups (e.g., host race evaluation). While “species” are typically used with ISA and groups are defined according to environmental characteristics, ISA could be used to estimate statistical associations between variables and any predefined group. We have provided some additional potentially relevant plant disease scenarios which ISA could be applied (Table 1) and note that a nested version of ISA exists (De Cáceres et al. 2010) where indicator values can be calculated for hierarchically organized groups. This nested ISA could be used to assess whether causal agents differ between tissue types or whether certain combination of management practices are associated with disease, for example.

RECOMMENDATIONS

Both published instances of ISA applications to plant pathosystems do not unambiguously demonstrate that ISA correctly identifies all disease causal agents, as formal pathogenicity studies on the organisms without statistically significant indicator values were not conducted for all species. ISA is a test for statistical association, and while it may correctly select causal agents, a statistically significant association does not indicate causation. Continued application will reveal ISA reliability as an analytical tool for plant disease study and in which situations it performs the best. ISA will be of limited use when a disease causal agent is either obvious or there are few potential pathogens to consider. However, plant pathologists may consider using ISA to statistically reevaluate previously studied diseases with potentially ambiguous causal agent determinations, such as the different races of *Fusarium* spp. Reanalysis may reveal additional, unrecognized causal agents, or interacting pathogens that form a disease complex, and act as a necessary evaluation of ISA's performance over a range of plant disease scenarios.

ACKNOWLEDGMENTS

We thank two anonymous reviewers, P. Ojiambo, and H. Scherm for comments that helped improve this manuscript.

LITERATURE CITED

- Aegerter, B. J., Gordon, T. R., and Davis, R. M. 2000. Occurrence and pathogenicity of fungi associated with melon root rot and vine decline in California. *Plant Dis.* 84:224-230.
- Arzanlou, M., Torbati, M., and Khodaei, S. 2013. Phenotypic and molecular characterization of *Plectosphaerella cucumerina* on bamboo from Iran. *Mycosphere* 4:647-651.
- Carlucci, A., Raimondo, M. L., Santos, J., and Phillips, A. J. L. 2012. *Plectosphaerella* species associated with root and collar rots of horticultural crops in southern Italy. *Personia* 28:34-48.
- Champaco, E. R., Martyn, R. D., and Miller, M. E. 1993. Comparison of *Fusarium solani* and *F. oxysporum* as causal agents of fruit rot and root rot of muskmelon. *HortSci.* 28:1174-1177.
- Chilosì, G., Reda, R., Aleandri, M. P., Camele, I., Altieri, L., Montuschi, C., Languasco, L., Rossi, V., Agosteo, G. E., Macrì, C., Carlucci, A., Lops, F., Mucci, M., Raimondo, M. L., and Frisullo, S. 2008. Fungi associated with root rot and collapse of melon in Italy. *EPPO Bull.* 38:147-154.
- De Cáceres, M., Legendre, P., and Moretti, M. 2010. Improving indicator species analysis by combining groups of sites. *Oikos* 119:1674-1684.
- de Gruyter, J., Woudenberg, J. H. C., Aveskamp, M. M., Verkley, G. J. M., Groenewald, J. Z., and Crous, P. W. 2010. Systematic reappraisal of species in *Phoma* section *Paraphoma*, *Pyrenopeziza* and *Pleurophoma*. *Mycologia* 102: 1066-1081.
- Dufrêne, M., and Legendre, P. 1997. Species assemblages and indicator species: The need for a flexible asymmetrical approach. *Ecol. Monogr.* 67:345-366.
- García-Jiménez, J., Armengol, J., Sales, R., Jordá, R., and Bruton, B. D. 2008. Fungal pathogens associated with melon collapse in Spain. *EPPO Bull.* 30:169-173.
- Gwynne, B. J., Gordon, T. R., and Davis, R. M. 1997. A new race of *Fusarium oxysporum* f. sp. *melonis* causing Fusarium wilt of muskmelon in the Central Valley of California. *Plant Dis.* 81:1095.
- Hill, M. O. 1979. TWINSPLAN-A FORTRAN program for arranging multivariate data in an ordered two-way table by classification of the individuals and attributes. Cornell University, Ithaca, NY.
- Hotelling, H. 1933. Analysis of a complex of statistical variables into principal components. *J. Exp. Psychol.* 24:417-441.
- Jagdale, G. B., Holladay, T., Brannen, P. M., Cline, W. O., Agudelo, P., Nyczepir, A. P., and Noe, J. P. 2013. Incidence and pathogenicity of plant-parasitic nematodes associated with blueberry (*Vaccinium* spp.) replant disease in Georgia and North Carolina. *J. Nematol.* 45:92-98.
- Kruskal, J. B. 1964. Nonmetric multidimensional scaling: A numerical method. *Psychometrika* 29:115-129.
- Martyn, R. D., and Miller, M. E. 1996. Monosporascus root rot and vine decline: An emerging disease of melons worldwide. *Plant Dis.* 80:716-725.
- McCune, B., and Grace, J. B. 2002. Analysis of ecological communities. MJM Software Design, Gleneden Beach, OR.
- Mehl, H. L., and Epstein, L. 2007. Identification of *Fusarium solani* f. sp. *cucurbitae* race 1 and race 2 with PCR and production of disease-free pumpkin seeds. *Plant Dis.* 91:1288-1292.
- Mielke, P. W., and Berry, K. J. 2001. Permutation Methods: A Distance Function Approach. Springer, New York.
- Rivedal, H. M. 2018. Diagnosis and etiology of a soilborne disease complex of winter squash (*Cucurbita maxima*) in the Willamette Valley, OR. Ph.D. dissertation. Oregon State University, Corvallis.
- Rivedal, H. M., Stone, A. G., Severns, P. M., and Johnson, K. 2020. Characterization of the fungal community associated with root, crown, and vascular symptoms in an undiagnosed yield decline of winter squash. *Phytobiomes J.* 4:178-192.
- Severns, P. M., Jagdale, G. B., Holladay, W., Brannen, P. M., Noe, J., and Cline, W. O. 2020. Potential for the pathogenicity of plant-parasitic nematodes associated with blueberries in Georgia and North Carolina. *Plant Health Prog.* 21:9-12.
- Siddig, A. H., Ellison, A. M., Ochs, A., Villar-Leeman, C., and Lau, M. K. 2016. How do ecologists select and use indicator species to monitor ecological change? Insights from 14 years of publication in ecological indicators. *Ecol. Indic.* 60:223-230.
- ter Braak, C. 1986. Canonical correspondence analysis: A new eigenvector technique for multivariate direct gradient analysis. *Ecology* 67:1167-1179.
- Zhang, N., O'Donnell, K., Sutton, D. A., Nalim, F. A., Summerbell, R. C., Padhye, A. A., and Geiser, D. M. 2006. Members of the *Fusarium solani* species complex that cause infections in both humans and plants are common in the environment. *J. Clin. Microbiol.* 44:2186-2190.
- Zuniga, T. L., Zitter, T. A., Gordon, T. R., Schroeder, D. T., and Okamoto, D. 1997. Characterization of pathogenic races of *Fusarium oxysporum* f. sp. *melonis* causing Fusarium wilt of melon in New York. *Plant Dis.* 81:592-596.