

Characterizing rice pest injury co-occurrence patterns at different irrigated lowland rice growing areas by network analysis

First Author^{1,*}, Co-Author² and Co-Author²

¹Laboratory X, Institute X, Department X, Organization X, City X, State XX (only USA, Canada and Australia), Country X

²Laboratory X, Institute X, Department X, Organization X, City X, State XX (only USA, Canada and Australia), Country X

Correspondence*:

Corresponding Author

Laboratory X, Institute X, Department X, Organization X, Street X, City X, State XX (only USA, Canada and Australia), Zip Code, X Country X, email@uni.edu

2 ABSTRACT

Survey data of 420 rice fields over three years across five countries, India, Indonesia, Philippines, Thailand, and Vietnam, were analysed using network analysis. The aim of the study was to characterize the patterns of pest injury co-occurrence in specific locations and to determine the key pests in the system. The results of the weighted co-occurrence networks indicated that ????. Therefore, network analysis provides insights into the pattern of co-occurrence which could be of particular interest for the identification of key factors with regard to different rice growing locations.

Keywords: Network analysis, Rice pests, Text Text Text Text Text

1 INTRODUCTION

Agricultural crop plants are frequently infected by more than one species of pests and pathogens at the same time. Many of these injuries may affect yield. Because of this co-occurrence in injuries, the idea of “crop health” has been highlighted and implemented to manage this combination of injuries or so-called injury profiles (Savary et al., 2006). Co-occurrence patterns of injury are beginning to provide important insight into these injury profiles, which possibly present co-occurring or anti-co-occurring relationships between injury-injury. Uncovering these patterns could have important implications in plant disease epidemiology and management. However, there are only a few reports of injuryinjury interactions in crop systems and the mechanisms of interactions are currently unknown and this could be a difficult task since complex patterns of injury profiles are related to environmental conditions, cultural practices, and geography (Wilocquet et al., 2008).

To address this issue, we used in-field surveys as a tool to develop ground-truth databases that allowed us to identify the major yield reducing pests in irrigated lowland rice agroecosystems. These sorts of databases provide an overview of the complex relationships between the crop, its management, pest injuries, and yields. Several previous studies Savary et al. (2000a,b); Dong et al. (2010); Reddy et al.

(2011) involved surveys that were used to identify relationships in an individual production situation, a set of factors including cultural practices, weather condition, socioeconomics, *etc.* that determine agricultural production, and the injury profiles using nonparametric multivariate analysis such as cluster analysis, correspondence analysis, or multiple correspondence analysis. Performing correspondence analysis (Savary et al., 1995), characterized the relationships between categorized levels of variables: actual yield, production situations, and injuries profiles. For example, stem rot and sheath blight are frequently found together with high mineral fertilizer inputs, low pesticide use, and good water management in transplanted rice crops, and overall their results led to the conclusions that observed injuries profiles were strongly associated with production situations and the level of actual yields.

We applied the technique from ecological study the co-occurrence analysis and network theory to reveal the patterns of co-occurrence of injury profiles. While existing statistic approach for analysis the survey data (e.g., cluster analysis, correspondence analysis, multiple correspondence analysis), the methods were present in this papers has the advantages of

Network theory is the study of relationships between entities (nodes) and connections between these entities (edges). Network theory has previously been used effectively to describe social and biological datasets, and it has been shown to be a useful tool for ???. Here, we consider pest injuries as nodes and create an edge between any two injuries if they are co-existing. We give an edge greater weight if the two injuries have strong co-occurrence at either end. The relationships will be more complex when the number of their components increased. A way to systemically model and intuitively interpret such relationships is the depiction as a graph or network. This approach has been widely used and proven very useful in biological studies (Moslonka-Lefebvre et al., 2011). Networks typically consist of nodes, usually representing components, while links between the nodes depict their interactions (Proulx et al., 2005). A correlation network is a type of network in which two nodes are connected if their respective correlation lies above a certain threshold. The construction of this network is obtained from pairwise correlation methods (Toubiana et al., 2013). By using appropriate correlation measure, correlation networks can capture biologically meaningful relationships, and discover valuable information in crop health surveys.

The aim of this study was to analyze how the structure of correlation network of pest injury co-occurrence differ across countries by quantifying the important aspects of the position of the specific pest injury, information based on network analysis could help to understand the formation and characteristics of co-occurrence patterns. Furthermore, key factors for could be identified with this additional information

Study sites, sampling and data collection

We used data from surveys conducted in Tamil Nadu and Kerala, India; West Java, Indonesia; Laguna, Philippines; Central Plain, Thailand; and Mekong river delta, Vietnam. These are major irrigated lowland rice growing areas where where rice is intensively cropped growing rice at least two seasons per year per year. Four-hundred twenty farmers' fields were sampled using a standardized protocol described by Savary and Castilla (2009) in, "A survey portfolio to characterize yield-reducing factors in rice", was used for data collection.

Injury variables were also simplified. Although a very large number of pathogens, insects, and weeds are harmful to rice, many are seldom considered to cause yield losses. Diseases such as narrow brown spot, bacterial leaf streak, leaf scald, and leaf smut, and insects such as rice bugs, rice hispa, and defoliators in general are not considered to represent major, widespread, yield-reducing factors. The study therefore concentrated on injuries listed in Table 1. A second aspect pertains to the injury mechanisms, and Table 1 includes injuries that were grouped in the field assessment procedure according to their nature: light stealers (BLB, BS, LB: proportion of injured leaves), senescence accelerators (BLB, SHB, LB: proportion of injured leaves, except for SHB), tissue users (leaves: RWM, LF: proportion of injured leaves; tillers: SR, SHB, DH: proportion of injured tillers; panicles: SHR, WH: proportion of injured panicles), assimilate sappers (PH: number of insects sampled), turgor reducers (at the tiller level: SR, SHB: proportion of injured tillers; at the panicle level: NB: proportion of injured panicles), and stand reducers (WA and WB).

Crop health survey data were collected in 420 farmers' fields from 2010 to 2012 for wet and dry seasons in different production environments across South and South East Asia. The variables collected included patterns of cropping practices, crop growth measurement and crop management status assessments, measurements of levels of injuries caused by pests, and direct measurements of actual yields from crop cuts. This data can be classified into three groups: cropping practices, crop injuries, and actual yield measurements.

Crop health survey data were collected through surveys comprising 420 farmers' fields from 2010 to 2012 for wet and dry seasons in different production environments across South and South East Asia. The survey protocol described in the IRRI publication, "A survey portfolio to characterize yield-reducing factors in rice", was used for data collection (Savary and Castilla, 2009). The variables collected included patterns of cropping practices, crop growth measurement and crop management status assessments, measurements of levels of injuries caused by pests, and direct measurements of actual yields from crop cuts. The data collected can be classified into three groups: cropping practices, injuries, and actual yield measurements.

Co-occurrence analysis

We considered co-occurrence both of positive and negative correlations based on Spearman's rank based correlation between pairs of pest injuries within each dataset with the strength of relationship (ρ from Spearman's correlation) represented by the correlation coefficient. The coefficients with p -values less than $p = 0.05$ were considered. Negative correlations (indicative of anti-co-occurrence relationship) were also included in analysis. The R function `cor.test` with parameter method 'Spearman' (package stats) was used for calculate Spearman's correlation coefficient (ρ), which is defined as the Pearson correlation coefficient between the ranked variables (R Core Team, 2015).

These correlation relationships were generated for each pair of injury within each location replicate as long as both injury had incidence value greater than 0. We made a network of co-occurrence relationships within each locations based on the strength of the correlation (ρ from the Spearman's correlation), and co-occurrence relationships were only included if they occurred across all locations. Though this method has been illustrated to produce some spurious co-occurrence relationships among data, this rank-based correlation statistic does not require any transformation of variables to fit assumptions of normality and may outperform Pearsons correlations. To increase our level of stringency that may reduce the appearance of spurious co-occurrences within our networks, pairwise relationships had to be consistent across all datasets of a given ecosystem type, greatly reducing the number of co-occurrence pairs.

Network analysis

Network models were used to illustrate the co-occurrence patterns of pest injuries within same locations, where injuries represent nodes and the presence of a co-occurrence relationship based on correlation is represented by an edge using the igraph package (Csardi and Nepusz, 2006), where each network was the union of positive or negative correlation coefficients (less than 0.25 or greater than 0.25) that were consistent within each location.

We were also interested in generating descriptive statistics about the network that may be important for understanding co-occurrence relationships. We produced network statistics that describe the position and connectedness of injuries within each co-occurrence network. Global network properties including the density, heterogeneity, centralization were computed by using `fundamentalNetworkConcepts` function from WGCNA package (Langfelder and Horvath, 2008) and for the basic properties such as number of nodes, edges can be computed by using functions from igraph package. Additionally, we also calculated the small-worldness index of the network by using `smallworldness` function qgraph package. For the node-wise properties including node degree, which is the number of co-occurrence relationships that an injury is involved in a network, we used the `degree` function from igraph package. We also calculated betweenness scores for each node (injury) using the `betweenness` function from igraph package, which is defined by the number of paths through a focal microbial node. Additionally, we calculated clustering coefficients, and eigenvector using the `transitivity` function for comparison to other networks.

Key factor analysis

Identification key factors in a network is very useful and widely used in social science. The way to identify key factors is to compare relative values of centrality such as eigenvector centrality and betweenness. Its apparent that many measures of centrality are correlated (Valente et al., 2008). The residual of linear relationship between eigenvector centrality and betweenness and regress betweenness on eigenvector centrality can used as an indicators of key factors. A node or factors with higher levels of betweenness and lower eigenvector centrality can be inferred that is central to the functioning of the network. Nodes with lower levels of betweenness and higher eigenvector centrality may be inferred that they are key to the functioning of the network.

2 RESULTS

We first visualized networks within each locations for both positive and negative co-occurrence relationships.

2.1 Rice pest injury co-occurrence networks

Table 1 shows the overall statistics of centralities for the identified modules. Interestingly, modules in clusters from healthy biofilms present lower centralization and higher density than the modules in the clusters from diseased biofilms, which may indicate that those modules could be more resilient to changes and the correlations among their members are high. Next step was to identify hubs in each of the modules. The question of which network elements are the most important cannot be answered unambiguously. Ranking nodes (species) in the network is accomplished by measuring different centrality indices using different algorithms. We used three different algorithms. First, we used degree centrality, which indicates the number of connections to other nodes in the network and has been used in numerous situations. For example, in the case of protein interactions, proteins with high degree centrality are more likely to be essential than those with low values of degree centrality [21]. Second, we utilized betweenness centrality, which indicates the relevance of a node as capable of holding together communicating nodes: the higher the value the higher the relevance of the node as an organizing regulatory node. The betweenness centrality of a node reflects the amount of control that this node exerts over the interactions of other nodes in the network [22]. Third, we used a double screening scheme (DSS), which combine two algorithms (Maximum Neighborhood Component and Density of Maximum Neighborhood Component) and has been shown to identify hubs that are missed by other algorithms [23]. In general, highly dense modules with low network centralization included many species, all of them with large number of species with high degree centralization and betweenness centrality (Table 1 and Table S1).

	country	Node	Link	diameter	avgConnect	avgGeodesic	density	smallworld	centralization	heterogeneity
1	PHL	21	33	1.85	0.28	2.26	0.07	0.82	0.19	
2	VNM	20	47	1.24	0.53	2.05	0.08	1.27	0.08	
3	THA	18	63	1.11	0.47	1.81	0.15	1.07	0.20	
4	IDN	22	48	1.40	0.46	2.25	0.06	1.24	0.07	
5	IND	10	17	1.38	0.55	2.02	0.14	1.71	0.14	

2.2 Indonesia

Brown (BS) has a high node strength, a relatively high betweenness, and a moderate clustering coefficient. Apparently, BS does play an important role in our sample of depressive and healthy persons: it can be activated very easily, since a lot of information flows through it (high betweenness), and, in turn, it can activate many other symptoms because it has many neighbors (high node strength, moderate clustering). Army worm (AW), leaf scald (LS), rice bug (RB), rat (RT), rice tungro disease (RTD), silver shoot (SS), and whitehead (WH) have a moderate strength and a very high clustering coefficient, but low betweenness. This indicates that these injuries in the network does not easily occur (low betweenness), but that, if they occurred, the cluster will possibly occur because of the high interconnectivity (high clustering coefficient). As opposed to other injuries, the injuries occurring on panicles (e.g., sheath blight (SHB), false smut (FSM), stink bug (STB), and deadheart (DH)) low scores on at least two centrality measures. Apparently, most panicle injuries either are less likely to occur simultaneously with other injuries (low clustering coefficient), or are less important for inducing other nodes through the network (low betweenness).

143 **2.3 India**

144 **2.4 Philippines**

145 **2.5 Thailand**

146 **2.6 Vietnam**

3 DISCUSSION

AUTHOR CONTRIBUTIONS

147 The statement about the authors and contributors can be up to several sentences long, describing the tasks
148 of individual authors referred to by their initials and should be included at the end of the manuscript before
149 the References section.

ACKNOWLEDGMENTS

150 Text
151 Text Text Text Text Text. Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
152 Text Text Text Text Text Text Text Text Text Text.

153 *Funding:* Text Text Text Text Text Text Text Text.

SUPPLEMENTAL DATA

154 Supplementary Material should be uploaded separately on submission, if there are Supplementary Figures,
155 please include the caption in the same file as the figure. LaTeX Supplementary Material templates can be
156 found in the Frontiers LaTeX folder

157 Text
158 Text Text Text Text Text.

REFERENCES

- 159 Csardi, G. and Nepusz, T. (2006). The igraph software package for complex network research.
160 *InterJournal Complex Systems*, 1695
- 161 Dong, K., Chen, B., Li, Z., Dong, Y., and Wang, H. (2010). A characterization of rice pests and
162 quantification of yield losses in the japonica rice zone of Yunnan China. *Crop Protection* 29, 603–611.
163 doi:10.1016/j.cropro.2010.01.007
- 164 Langfelder, P. and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis.
165 *BMC Bioinformatics* 9, 559. doi:10.1186/1471-2105-9-559
- 166 Moslonka-Lefebvre, M., Finley, A., Dorigatti, I., Dehnen-Schmutz, K., Harwood, T., Jeger, M. J., et al.
167 (2011). Networks in Plant Epidemiology: From Genes to Landscapes Countries, and Continents.
168 *Phytopathology* 101, 392–403. doi:10.1094/phyto-07-10-0192
- 169 Proulx, S., Promislow, D., and Phillips, P. (2005). Network thinking in ecology and evolution. *Trends in*
170 *Ecology & Evolution* 20, 345–353. doi:10.1016/j.tree.2005.04.004
- 171 R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for
172 Statistical Computing, Vienna, Austria

- 173 Reddy, C., Laha, G., Prasad, M., Krishnaveni, D., Castilla, N., Nelson, A., et al. (2011). Characterizing
 174 multiple linkages between individual diseases crop health syndromes, germplasm deployment, and rice
 175 production situations in India. *Field Crops Research* 120, 241–253. doi:10.1016/j.fcr.2010.10.005
- 176 Savary, S. and Castilla, N. (2009). A survey portfolio to characterize yield-reducing factors in rice. *IRRI*
 177 *Discussion Paper No 18*, 32
- 178 Savary, S., Madden, L., Zadoks, J., and Klein-Gebbinck, H. (1995). Use of Categorical Information and
 179 Correspondence Analysis in Plant Disease Epidemiology. In *Advances in Botanical Research* (Elsevier
 180 BV). 213–240. doi:10.1016/s0065-2296(08)60014-2
- 181 Savary, S., Teng, P. S., Willocquet, L., and Nutter, F. W. (2006). Quantification and Modeling of Crop
 182 Losses: A Review of Purposes. *Annu. Rev. Phytopathol.* 44, 89–112. doi:10.1146/annurev.phyto.44.
 183 070505.143342
- 184 Savary, S., Willocquet, L., Elazegui, F. A., Castilla, N. P., and Teng, P. S. (2000a). Rice Pest Constraints
 185 in Tropical Asia: Quantification of Yield Losses Due to Rice Pests in a Range of Production Situations.
 186 *Plant Disease* 84, 357–369. doi:10.1094/pdis.2000.84.3.357
- 187 Savary, S., Willocquet, L., Elazegui, F. A., Teng, P. S., Du, P. V., Zhu, D., et al. (2000b). Rice Pest
 188 Constraints in Tropical Asia: Characterization of Injury Profiles in Relation to Production Situations.
 189 *Plant Disease* 84, 341–356. doi:10.1094/pdis.2000.84.3.341
- 190 Toubiana, D., Fernie, A. R., Nikoloski, Z., and Fait, A. (2013). Network analysis: tackling complex data
 191 to study plant metabolism. *Trends in Biotechnology* 31, 29–36. doi:10.1016/j.tibtech.2012.10.011
- 192 Valente, T. W., Coronges, K., Lakon, C., and Costenbader, E. (2008). How Correlated Are Network
 193 Centrality Measures? *Connections (Toronto, Ont.)* 28, 16–26
- 194 Willocquet, L., Aubertot, J., Lebard, S., Robert, C., Lannou, C., and Savary, S. (2008). Simulating
 195 multiple pest damage in varying winter wheat production situations. *Field Crops Research* 107, 12–28.
 196 doi:10.1016/j.fcr.2007.12.013

FIGURES

