\chapter{Differential networks reveal the dynamics of animal pests and disease co-occurrences}

\subsection{Introduction}

Rice textit{Oryza sativa}) is a major crop in South and Southeast Asian. Generally, rice farmers cultivate 2 rice crops per year, with the typical seasonal crop cycles or rotations being rice-rice-fallow or rice-rice-secondary crops (corn, soybean, peanut). The Food and Agriculture Organization of the United Nations (FAO) estimates that approximately 70 percent of total lowland rice area produces 2 rice crops each year. is cultivated in the wet season, while another is in the dry season. The important role of seasonal cropping in the temporal dynamics of animal pests and diseases has been studied under farmers field survey in South and Southeast Asia by the use of multivariate techniques \citet{Savary\_2000\_Characterization, Willocquet\_2008\_Simulating}. The previous studies showed that Injuries profiles (the combination of injuries) differ from season to season in term of weather pattern. In the dry season, crop losses were lower than in the wet season. A previous study based on surveys done in farmers’ rice fields in the region of lowland rice were shown to be strongly associated with injury profiles.

In the previous chapter, the co-occurrence networks to yield co-occurrence networks, a methodological approach which has already proved fruitful in a variety of different applications. Plant injuries caused by pests maybe affect yield production. Therefore, in this chapter, I attempted to characterize the patterns of rice injuries by studying the changes in the co-occurrence patterns of rice injuries (textit{e.g} disease incidence, animal pest injury incidence) at different yield levels.

%Conceptually, network comparison is the process of contrasting two or more interaction networks, representing different species, conditions, interaction types or time points. This process aims to answer a number of fundamental biological questions: which proteins, protein interac- tions and groups of interactions are likely to have equivalent functions across species? Based on these similarities, can we predict new functional information about proteins and interactions that are poorly character- ized? What do these relationships tell us about the evolution of proteins, networks and whole species?

Differential network analysis aims to compare the connectivity of two nodes at 2 different conditions. As demonstrated by several studies, differential networks can identify important nodes implicated in my fields, and also provide critical novel insights not obtainable using other approaches. In this work, I explore the the properties of network of a complex association of rice injuries at different yield levels. Elucidating the rice injuries association represents a key challenge, not only for achieving a deeper understanding of injury association (injury profiles) but also for identifying the unique association. Given that the injury association is governed by a complex network of injuries association, it seems natural to explore network properties which may help elucidate some of different association presenting in the different seasons.

My overall objective was to explored the relation between local differential association and discuss the meaning of the results in the context of the crop health management, and discuss the potential implications of our findings for development of pest management with a view to future studies.

From resulted networks from the previous chapter, it is found that co-occurrence networks of rice injuries are not static; rather, they exhibit dynamic adaptations in response to both seasonal changes and locations. Influences that determine network context can be divided into four categories. (1) seasonal weather strongly determines network structures. (2) location change. Thus overall is to determine the

Genetic background strongly determines network behavior and gives rise to significant differences across individuals (and even cells in the special case of cancer). (2) Cell lineages have dramatically different network structures because of epigenetic changes and differential expression of genes. (3) Tissue milieu can reprogram networks and their behaviors, as stromal cells do for tumors. (4) Exogenous signals, such as nutrients and other chemicals, affect networks (Figure 1). Ultimately, health or disease emerges from an individual’s integration of internal and external cues.

Overall objective is to

evaluate the impact of adapting the SimCast model to predict the risk of late blight based on daily and monthly weather data. My first objective for this project was to develop disease prediction models based on daily and monthly weather means and compare to results based on hourly weather data. The second objective was to compare the blight unit predictions of models constructed from weather data sets specific to potato growing regions with models constructed with a data set that represents a broad range climate types. The third objective was to compare late blight risk predictions based on hourly, daily, and monthly weather averages to observed late blight severity data sets from four countries.

In summary, our study provides a mathematical/bio- informatic framework for network construction based on metagenomics data such as sequencing [28] and microarray hybridization data [27]. It is useful, as demonstrated with the microbial communities under ex- perimental warming, for dissecting interactions within a microbial community as well as with environment, thus

Figure 1 shows an overview of our approach. For each gene in the network, its differential expression level is determined by transcriptome-wide microarray experiments of mutant vs. wild- type lines. Then differentially expressed neighborhoods in this network are considered for all candidate genes from a chromo- somal region of interest (e.g., identified in a linkage study). Candidate genes with genes in their neighborhood having highly differentially expressed levels are strong candidates. The neigh- borhood of a candidate gene is defined by direct connections in the global distance network, whereby its size can vary. The smaller the distance from a neighboring gene to the candidate, the closer the neighboring gene is in the network. We have chosen to work with the notion of a soft neighborhood by which we mean that the neighborhood of a gene is not a limited set of gene, but rather a weighting function that decreases monotonically as a function of the distance from the gene, but potentially covers the whole network (this idea is reminiscent of the idea of fuzzy membership in fuzzy c-means clustering).

\subsection{Materials and Methods}

The first objective was to develop . To do this, I compared

The second objective was to compare the

The third objective was to compare SimCast Daily Means and SimCast Monthly Means output to disease severity observations from several countries. For the third objective, late blight severity and hourly weather data from 19 cultivar, site-year combinations

I used survey data were collected from farmers' fields ,analyzed co-occurrence relationships of rice injuries, and performed network analysis as described previously. Actual yield estimates were collected from each farmers' fields surveyed. Before analyzing. Three yield levels were I grouped the survey data into 6 groups by different yield levels and seasons.

\textbf{Network Comparison Measures}

\begin{equation}

A\_{diff} = \left\{\begin{matrix}

1 & \text{when } C\_{xy}^D > C\_{xy}^W \text{ at } P\_{z\_{xy}} \text{-value} < 0.05 \\

0 & \text{when } P\_{z\_{xy}} \text{-value} > 0.05 \\

-1 & \text{when } C\_{xy}^W > C\_{xy}^D \text{ at } P\_{z\_{xy}} \text{-value} < 0.05

\end{matrix}\right

\end{equation}

For the pair of $x\_{i}$ injury and $y\_{i}$ variable, I denoted the correlation coefficient based on Spearman's correlation coefficient by $c\_{xy}^1$ and $c\_{xy}^2$ in networks 1 and 2, respectively. To test whether the 2 correlation coefficients were significantly different of each network, I used Spearman's correlation coefficients, the p-values of the correlation test, the difference of the 2 correlations, the corresponding $p$-values, and the result of Fisher's z-test. First, I applied the Fisher $z$-transformation in order to stabilize variances due to sample size.

\begin{equation}

z\_{xy} = \frac{1}{2} \log\left[{\frac{1 + c\_{xy}}{1 - c\_{xy}}}\right]

\end{equation}

if we let $z\_{xy}^D $ and $z\_{xy}^W$ denote the $z$- transformation for dry and wet season variable pairs, respectively.

Next, differences between the two correlations can be test using following.

\begin{equation}

P(Z\geq \left | \frac{z\_{xy}^D - z\_{xy}^W}{\sqrt{\frac{1}{N\_{D}-3}+ \frac{1}{N\_{W}-3}}} \right |

\end{equation}

$N\_{D}$ and $N\_{W}$ represent the sample size for each of season network for each country. The $Z$ has an approximately Gaussian distribution under null hypotheses that the population correlations are equal.

In the differential injury set coexpression test, we obtained

parametric p values and selected significant results according

to the threshold determined from the p-value distribution.

The selected pathway pairs were retested in a

nonparametric manner.

The pair-wise correlation significant are considered at $p$vlaue < 0.05.

In the above equations, CC indicates the correlation coefficient of single gene pair. Zf1 and Zf2 are the Fisher's Z-transformed correlation coefficients of conditions 1 and 2. N1 and N2 are the number of samples in conditions 1 and 2, respectively. From the normal distribution, $p$-values for differential coexpression tests were obtained according to the difference between the Z values.

During calculation, three p values were obtained for each gene pair. The p values were those of correlation coefficients from condition 1 and condition 2, and from the difference between Fisher's Z-transformed correlation coefficients.

The adjacency matrix $A\_{diff}$ of this network formally expresses injury occurrences, and is written in $A=[Z\_{ij}]$, which is

\begin{equation}

C\_{ij} = \begin{cases}

C\_{ij} & \text{if } \rho > 0 \text{ and } p\text{-value } < 0.05 \\

0 & \text{otherwise}

\end{cases}

\end{equation}

Differ

Statistical analysis

Statistical analysis was performed using SAS® statistical software package (SAS® Institute Inc 2008). Due to the small sample size in each pen and the distribution of the data, a Kruskal-Wallis-Test was used to assess the differences between the centrality parameters across the three observed age levels. Furthermore, a Spearman rank correlation was performed to characterize the relations between the centrality parameters.

Generally, correlation or differential correlation does not necessarily reflect causal relationships (Markowetz and Spang, 2007; Steuer, 2006). Although they should include causal relationships, it should be noted that correlation networks based on undirected graphs cannot easily distinguish

o assess the significance of differential coexpression using the dispersion statistics is presented i

Conclusions

The R package DiffCorr affords users a simple and effective framework to detect differential correlations between 2 conditions in omics data. The package is based on Fisher's z-test and is simple to calculate differential correlations. The approach is useful for the first step towards inferring causal relationships and detecting biomarker candidates. The DiffCorr base on the concept of “differential network biology” (de la Fuente, 2010; Ideker and Krogan, 2012) is suitable not only for transcriptomic and metabolomic data, but also for proteomic data, genome-wide asso- ciation studies and integrated omics data (Fukushima et al., 2009; Kim et al., 2010).