OMICS Group Conferences
Accelerating Scientific Discovery

2nd International Conference on
Proteomics & Bioinformatics

July 2-4, 2012 Embassy Suites Las Vegas, USA

Resistance protein network obtained through genomic data integration

Luis Leal[1], Liliana López-Kleine[1], Camilo López[2] and Alvaro Pérez[2]
[1]Statistics Department, Universidad Nacional de Colombia, Colombia
[2]Biology Department, Universidad Nacional de Colombia, Colombia

A huge quantity of high-throughput data on several plants is produced and used for knowledge extraction daily. Systemic approaches using either genomic or post-genomic data have been done before for understanding of pathogen responses and biochemical processes of plants, especially in Arabidopsis thaliana. These and other preceding work has enabled the development of disease control strategies. Nevertheless, the precise role of many genes has not been explained and an important issue remains to be addressed: To construct resistance protein networks integrating as much available information as possible. The biological protein interaction networks are useful to represent both genomic and post-genomic data, allowing us for the elaboration of biological hypothesis to be validated in wet lab experiments.

Since the knowledge of A. thaliana is more extensive, a resistance protein network based on this plant is a good point of reference to validate our method. On the other hand, a network for a fairly unknown and economically important plant as cassava (Manihot esculenta), which genome was recently sequenced, brings new possibilities to expand our understanding about the specie's response to pathogen attacks.

We collected all available genomic information for A. thaliana and cassava from several databases, this includes presence of conserved domains on resistance proteins, predictions on cellular localization, PFAM, KEGG, GO, miRNA targets, etc. and available post-genomic information microarray and RNA-sequencing data related to resistance experiments.

Initially, some classical descriptive multivariate analysis such as Multiple Correspondence Analysis, Hierarchical and Non-hierarchical Cluster Analysis were conducted on categorical data, as a first step to evaluate the information contained and observe the behavior of genes potentially implicated in defense processes. We suggested protein function for genes with unknown function if they are statistically similar to genes for which the protein product function is already known.

Subsequently, all data types were represented as kernels and a kernel method allowing the determination of distances between genes, based on all genomic data available, was applied to infer partners of potential resistance genes.

Here we present our approach for obtaining, integrating and analyzing several types of genomic data in order to construct a protein network representing interactions of resistance proteins. Our first results on relationship of resistance genes with other candidates to participate in defense processes will be discussed.

Biography

Luis Leal is a Chemical Engineer and has completed his Statistics Specialization at the age of 24 from the Universidad Nacional de Colombia. He is studying a Master of Science in Statistics/Biostatistics. He is member of the Methods in Biostatistics group at the Statistics Department of the same university where he works as biostatistician. Together with the biologist Alvaro Perez and the leaders PhD. Biol. Liliana López-Kleine and PhD. Biol. Camilo López, he is working in a common project on cassava protein network reconstruction.

lgleala@unal.edu.co