

VirtualPlant: A software platform to support Systems Biology research in the post-genomic era.

Katari¹ MS, Nowicki¹ SD, Aceituno, F.², Poultney³ C., Nero, D¹., Chiaromonte³, F., Shasha³ D.E., Coruzzi¹ G.M., and **Gutiérrez^{1,2}, R.A.**

¹*Department of Biology, New York University.*

²*Departamento de Genética Molecular y Microbiología. P. Universidad Católica de Chile.*

³*Courant Institute of Mathematical Sciences. New York University.*

In the post-genomic era, data generation is no longer the limiting factor for advancing biological research. Instead, data integration, analysis and interpretation are the bottlenecks and challenges that biologists face everyday in genomic research. We wish to aid biologists to take advantage of the burgeoning amount of genomic data, by developing a software platform that enables scientists to visualize, integrate and analyze genomic data from a systems biology perspective. We term this software platform “VirtualPlant”.

The VirtualPlant platform integrates genome-wide data concerning the known and predicted relationships among genes, proteins and molecules, as well as genome-scale experimental measurements. VirtualPlant goes beyond data integration to conceptual synthesis, by using novel visualization techniques to render the multivariate information in visual formats that facilitate extraction of biological concepts. VirtualPlant also automates the use of mathematical and statistical methods to help summarize and quantify the data. VirtualPlant implements and combines these quantitative and qualitative (e.g. visual) approaches to data integration and analysis using a platform-independent, user-friendly, web-accessible interface. VirtualPlant can be used to help biologists mine genomic data to address grand challenge questions in plant biology such as: “What are the molecular mechanisms by which internal and external perturbations affect processes and gene networks controlling growth and development?” As a proof of principle, we will discuss how VirtualPlant was used to identify gene networks and regulatory hubs that control N-responsive gene networks.

Whereas the VirtualPlant project was developed specifically for Arabidopsis, the data structures, algorithms, and visualization tools are designed in a species-independent way. Thus, the informatic, math, statistic and visualization tools that we have developed for VirtualPlant can be used to perform systems biological analysis of the cellular and physiological responses of any organism for which genomic data is available. The VirtualPlant system is available from <http://www.virtualplant.org>