



BIOTECHNOLOGY RESEARCH FOR A COMPLEX WORLD

20th anniversary of the EC-US Task Force on Biotechnology Research

2 June 2010



A SYSTEMS APPROACH IDENTIFIES REGULATORY NETWORKS AFFECTING PLANT GROWTH AND DEVELOPMENT

Katari M¹, Gutierrez R², Shasha D³, Krouk G¹, Lee E⁴, Kolokotronis S⁴, Cibrian A⁵, Chiu J, Stevenson D⁵, McCombie WR⁶, Martienssen R⁶, DeSalle R⁴ & Coruzzi G¹

Center for Genomics & Systems Biology, NYU, New York, NY¹;

Catolica Universidad de Chile, Santiago, Chile²

Courant Institute of Mathematical Sciences, NYU, New York, NY³;

Sackler Institute for Comparative Genomics, American Museum of Natural History, NY⁴;

New York Botanical Garden, Bronx, NY⁵;

Cold Spring Harbor Labs, Cold Spring Harbor, NY⁶

A goal of Systems Biology is to explain how gene network changes evoke whole system responses. Importantly, systems approaches can be used to generate predictive models that can predict how network responses change under untested conditions or in response to gene modifications. Such predictive modeling of the system will enable in silico testing of gene modifications with important implications for agriculture (1). As a first step towards this ambitious goal, we created an Arabidopsis “multinetwork” that contains all known gene interactions (2). In this “mutlinetwork” the “edges” connecting Arabidopsis gene “nodes” are supported by multiple data/evidence including metabolic pathway connections, protein:protein and protein: DNA interactions, microarray data, microRNA:target datasets, and literature-based interactions (~10,000 nodes and 230,000 interactions) (www.virtualplant.org) (2). By querying this Arabidopsis multinetwork with microarray data from organs and specific cell-types, the resulting subnetworks enabled us to derive new, testable hypotheses for mechanisms controlling nitrogen sensing and signaling in plants. Validated examples of the success of this approach include the discovery of a role for the myb transcription factor CCA1 (a central clock gene) which acts as a network hub that mediates organic-N sensing (3). A second example is the discovery of the role of a microRNA-TF motif involved in the N-regulation of lateral root outgrowth (4). To next implement dynamic and predictive networks, we generated a High Resolution Dynamic Transcriptome (HRDT) which detected new gene sets (including TFs and kinases) induced as early as 3 min of NO₃⁻-treatment. By analyzing this time-series data with a machine learning approach called “state-space” modeling, we were able to accurately predict the change in direction of gene expression at untested states with >80% accuracy (5). As a first step to adapting such systems approaches to comparative genomics, we have exploited the fully sequenced genomes to automate genome-scale parsimony-based ortholog determination (OrthologID) (<http://nypg.bio.nyu.edu/orthologid>) (6). We next used these orthologs to construct phylogenetic trees consisting of thousands of genes/species within a parsimony framework (7). By automating this phylogenomic pipeline, we assembled a matrix of 22,833 orthologs from all available plant genome sequences, including all fully sequenced genomes and species with >2,000 EST/unigenes. This gene matrix called *BIGPLANTv1*, was used to build a robust phylogenomic tree of based on all available plant genome sequences (>150 taxa) (<http://nypg.bio.nyu.edu/orthologid/bigplant>) (8). This parsimony framework can now be used to identify the genes and overrepresented GO terms that support specific nodes or clades, thus enabling a functional phylogenomic approach for discovery of genes involved in the evolution of important agronomic and other traits.

This research is supported by NSF Plant Genome, DBI and 2010 Grants, NIH and DOE. *References:* (1) Gutierrez et al (2005) *Plant Physiol.* 138: 550; (2) Gutiérrez et al (2007) *Genome Biology*, 8: R7 ; (3) Gutiérrez et al (2008) *Proc. Natl. Acad. Sci. USA*, 105: 4939; (4) Gifford et al (2008) *Proc. Natl. Acad. Sci. USA*, 105: 803.; (5) Krouk et al (submitted), (6) Chiu J, et al (2006) *Bioinformatics*, 22: 699, (7) delà Torre et al (2010) *PLoS One*, v6, 5764, (8) Lee et al (Submitted) .