**PROJECT SUMMARY** **ABI Sustaining:** “**VirtualPlant: A Software Platform For Systems Biology**".

**1. List of senior personnel**. (PI, co-PIs, Key collaborators) with home institution

**PI**: Gloria Coruzzi, NYU Center for Genomics & Systems Biology

**Co-PIs**: Dennis Shasha, NYU Courant Institute of Mathematical Sciences,

Manpreet Katari, NYU Biology, Center for Genomics & Systems Biology.

**Senior Personnel**: Kranthi Varala, NYU Center for Genomics & Systems Biology

**2. Intellectual merit of the proposed activity.**  Our vision is to enable Systems Biology studies across plant species by sustaining our VirtualPlant project. The VirtualPlant software platform ([www.virtualplant.org](http://www.virtualplant.org)) was developed under an NSF Arabidopsis 2010 Grant (DBI-0445666) and includes tools for data analysis, integration and visualization such as the Arabidopsis multinetwork data, expression and network analysis, and data integration tools [[1](#_ENREF_1)]. VirtualPlant has played an integral part in enabling plant systems biology research in many plant labs around the world. Our recent expansion of VirtualPlant to include important crops, such as Rice and Maize, will allow the platform to reach a much wider community of plant biologists, and will, through comparisons to Arabidopsis, support translational research through this sustaining grant. For example, VirtualPlant can allow researchers to predict how an interacting network of genes/products in crop genomes will react *as a system* in response to an environmental change or genetic modifications. As demonstrated in the case study (Section 2), we are able to achieve this ambitious goal by using VirtualPlant for Maize based on data from a maize experiment, the Arabidopsis multinetwork, and homology. The results of such analyses generate a set of testable hypotheses across species. Creation of a VirtualPlant database and interface for a new species is now semi-automated, allowing us to extend VirtualPlant to many more species such as *Medicago truncatula* and *Glycine max,* and other crop genomes as they come on line in the near future. **We divide the work into three sections:**

**Section 1: Scientific value of Virtual Plant to enabling Systems Biology studies within the plant genomics community.** Virtual Plant enables seamless integration of data and tools into a single operating environment. The “Gene Cart” enables researchers to store results, enabling iterative cycles of analysis, a highlight of Systems Biology research. VP’s Biologist Friendly GUIs enable plant biologists, with no computer training to analyze their genomic data providing biological insights. Since its publication in Feb 2010, VirtualPlant and its tools have been cited more than 50 times. In Section 1, we demonstrate how the community is using VirtualPlant and highlight the diverse profile of our user base. VP currently has 1,033 registered users from 42 different countries spanning academia and industry.

**Section 2. VirtualPlant: Enabling Translational Research**. Section 2 describes a case study of how a Maize researcher can use VirtualPlant to build a hypothesis regarding nitrogen-responsive gene networks in Maize partly by transferring knowledge from Arabidopsis to Maize. With the advent of Next-gen sequencing, we expect an increase in the amount of experimental data from crop species. We already have VirtualPlant operational for Arabidopsis, Maize and Rice. Part of this sustaining grant will be to update and maintain the data and annotations. Since we have semi-automated the induction of new species from Phytozome, VirtualPlant can readily support crop genomes including Soy, and other important genomes such as Brassica, Medicago, Grape, Tomato, Sorghum, and Poplar.

**Section 3: VirtualPlant: Dissemination and User support/interaction**: VirtualPlant offers a set of systems biology research tools that are complementary to the excellent resources found on Maizesequence and Soybase. We will work with these groups to enable seamless communication between these databases and VirtualPlant. Further, we will attend the Maize and Soy annual meetings and conduct VP workshops to support and assess the needs of the crop user communities. Finally, we will provide free Webinars and video tutorials hosted on the VirtualPlant site to introduce new users to the interface and tools available. We will also continue to support our existing user base through the help section on our website and a mailing list where users can ask questions.

**Justification for ABI Goals**: Sustaining Virtual Plant will meet the ABI goals of supporting the analysis and dissemination of biological knowledge through informatics tools and plant systems biology resources that support transformative research in the biological community.

**3. Broader impacts of the proposed research.** This project is the result of a long-standing and highly successful collaboration between biologists at NYU and elsewhere, and computer scientists at NYU's Courant Institute of Mathematical Sciences. The Systems Biology tools resulting from this project are written in a species-independent manner so that they can empower biologists to use genomic data to predict a spectrum of gene networks in biology with broad applications to agriculture, the environment, and human health. In addition to scientific results, this collaboration extends to joint training of biologists and computer scientists in the field of Systems Biology.

**RESULTS FROM PRIOR NSF SUPPORT:** This NSF ABI Sustaining grant is most closely related to and leverages an NSF DBI Grant entitled “Conceptual Data Integration for the Virtual Plant” (NSF Database Activities: DBI-0445666). The VirtualPlant software platform (www.virtualplant.org) [[1](#_ENREF_1)] developed in that parent grant integrates genome-wide data from Arabidopsis concerning the known and predicted relationships among genes, proteins and molecules, as well as genome-scale experimental measurements. VirtualPlant also provides tools that render multivariate information into integrated visual displays including networks, directed acyclic graphs, and other novel visualization formats to highlight biological implications. Tools embodied in the VirtualPlant system have been used by us [[1-4](#_ENREF_1)] and other notable Arabidopsis research teams [[5-9](#_ENREF_5)] to generate hypotheses in Arabidopsis that were subsequently experimentally validated.

***Our parent NSF VirtualPlant grant had three major goals***: **Integration & Prediction**, **Visualization**, and **Synthesis**, which we have accomplished, as outlined below.

**Aim 1.** **Data** **Integration & Prediction**: ***The Arabidopsis Multinetwork*.** Our VirtualPlant project included assembling the first multinetwork for Arabidopsis, a first step towards a molecular wiring diagram of the plant cell [[1](#_ENREF_1),[2](#_ENREF_2)]. The Arabidopsis multinetwork in VirtualPlant has 16,562 nodes (of which 13,960 are genes) and 97,423 interactions (Fig. XB, & Table X). The multinetwork enables researchers to interpret transcriptome data in the context of all known sources of interaction including protein, DNA, RNA, etc. The multinetwork has been used to integrate plant gene/protein interaction data from a number of sources including protein:DNA interaction data [[5](#_ENREF_5)], protein:protein data [[10](#_ENREF_10)], and more recently the Arabidopsis interactome data [[11](#_ENREF_11)]. In one example, a query against the Arabidopsis multinetwork with 834 nitrogen-regulated genes resulted in a sub-network of 369 genes connected by one (or more) “expression correlation edges” [[3](#_ENREF_3)]. At the top of the resulting list of network TF “hubs” (with 47 connections to targets in the N-regulatory network) was the central clock control gene CCA1, a Myb family transcription factor (TF) [[3](#_ENREF_3)]. This discovery enabled us to derive and validate the novel hypothesis that nitrogen-regulation of CCA1 mRNA expression sets the circadian clock. Other examples of hypotheses derived and validated using the VirtualPlant multinetwork are reported in [[2](#_ENREF_2),[3](#_ENREF_3),[12](#_ENREF_12)].

**Sibohan Brady (UC Davis) re multinetwor**k: “I work on gene regulatory networks in plants, and have published several recent papers that comprise the mapping and analysis of these networks [[5](#_ENREF_5),[13](#_ENREF_13)]. A requirement for publication was to have all protein- DNA interactions hosted in a public database for free download. None of the standard interaction databases would accept our interactions since they were in yeast-one-hybrid or genetic interaction format. Instead, Virtual Plant was one of two plant-based websites that accepted this data (the other was AGRIS). It was very easy to work with their administrator to have our data uploaded in the multinetwork, and I am very grateful that they provide this resource in addition to many other *fantastic tools* that my lab uses regularly.”

**Aims 2 & 3. Synthesis and Visualization: *VirtualPlant’s primary analysis tools and functions.*** In addition to the Arabidopsis multinetwork, the VirtualPlant platform houses other Systems Biology tools for data analysis, integration and visualization. Below is a list of some of the most popular tools deployed through VirtualPlant.

**The "GeneCart" Function**: A key challenge to analyzing genomic data is the complex analysis workflow required by currently available software. VirtualPlant solves this problem by integrating multiple tools into a single platform that standardizes the representation of their inputs and outputs so that the output of almost any analysis can be stored in the “GeneCart” and later serve as the input to any VirtualPlant analysis tool. The GeneCart can store lists of genes, microarray and RNA-seq experiments, and gene networks. An item or items can be used as input for an analysis tool and the output of the tool can be stored in the GeneCart. This feature facilitates Systems Biology's iterative cycles of data analysis and experimentation [[14](#_ENREF_14),[15](#_ENREF_15)]. Three working examples described in [[1](#_ENREF_1)], illustrate how VirtualPlant can be used to perform iterative data analyses that build and refine testable biological hypotheses.

**Set Operations.** VirtualPlant supports the set operations union, intersect, and symmetric difference. In addition to the combinatorial operations, sets can also be renamed, deleted, or assigned to a “group”. Groups, functioning like folders in a file system, can be arbitrarily deeply nested and are used to organize gene sets. This is fundamental but important feature as one frequent user **Mary Wildermuth** (U.C Berkeley) notes: “[Dennis thinks we should eliminate the part that is yellow; may be true but too weak] *It is surprisingly difficult to get a simple output of the intersection of two gene lists using other available resources*, *VirtualPlant rapidly allows one to find intersections, combine lists (removing redundant genes), and store the output for further analysis.”*

**BioMaps (Fig. X)**: BioMaps takes one or more sets of genes and determines which functional terms (GO or MIPS are statistically over-represented in each set, with respect to a background population (e.g. Arabidopsis genome). The output is presented in either a tabular format that can be downloaded to Microsoft Excel or in an interactive graphical representation of the GO tree and the significant terms. Users can zoom the GO Directed Acyclic Graph (DAG) to any level of detail, and select GO terms. They can also query the GO terms and use union, intersect, and difference set operations to form new genelists and send them to the cart. Also, they can adjust a slider that sets a p-value threshold to which hides GO terms whose p-values exceed the threshold. A png image of the BioMaps DAG can be saved. **Mary Wildermuth (UC Berkeley) writes**: ‘*BioMaps in VirtualPlant is a fabulous, versatile, and functional tool that we routinely use to determine functional processes statistically enriched in our gene sets.* *The appropriate control set can be readily selected to ensure proper statistical treatment, both MIPS and TAIR classifications are available for use, and the output can be obtained in a number of ways (e.g. as lists and graphical networks).  Because the output is not just the enriched category with the statistics, but also the actual genes in each category, additional functional category-based analysis is made easy.  We cite the use of BioMaps in VirtualPlant in four of the publications [*[*6-9*](#_ENREF_6)*].  In these papers, we used BioMaps to ascertain the most significant functional processes (and the specific genes within those processes) impacted by parasite, chemical treatment, and/or disruption of a gene.  We then focus in on identified processes to further experimentally assess the role of the process or gene in the process in our system”.*

**Sungear**: Sungear is a visually interactive and biologist-driven exploration of comparisons of the results of many experiments on a genomic scale. Sungear can represent an arbitrary number of experiments/lists, all of their disjoint intersections, and their related ontological terms. The position of a circle and arrows emanating from it, indicate the input lists of which it is a subset. The size of a circle is proportional to the number of genes in the intersection of those lists (see [[16](#_ENREF_16)]). Many biologists find Sungear to be an extremely powerful and interactive tool for analyzing the interrelationships between sets of genes [[17](#_ENREF_17)]. **Mary Wildermuth** writes: “*Sungear allows for the easy comparison of more than two genelists at one time - to visualize where overlap occurs and does not occur. Though the idea sounds simple, implementation and visualization of such an output is not and VirtualPlant's Sungear is a unique tool. We use it to provide an integrated view of factors of interest. For example, it helps us prioritize sets of mutants to analyze in concert.”*

**GeneSect:** Genesect is a non-parametric randomization test to determine whether the overlap between two gene lists is higher/lower than expected by chance. The test consists of 1,000 tests, each of which randomly and uniformly selects two lists, out of the user-specified background population, where the lists are the same size as the observed lists. Thus, the algorithm counts the number (n) of times that the intersection size of the random lists is equal or greater/less than the intersection observed for the two tested gene lists. A p-value is thus generated equal to n/1,000.

**Microarray and RNA-seq Analysis Tools:** Transcriptome analysis is an important part of systems biology today. One of the main advantages of RNA-seq is that it does not require knowledge of the reference genome. For this purpose we expect a higher influx of transcriptome data in the near future. To accommodate transcriptome data we allow users to upload both their microarray data (CEL files or normalized matrix) and RNA-seq, data represented by a matrix containing the gene name and the number of reads that mapped to the genes. For microarray analysis we have a set of common statistical methods for determining probes that are **differentially expressed genes** including, T-test, Log Ratio Cutoff, RankProduct, and ANOVA. For RNA-seq, we use Deseq to determine genes that are differentially expressed. In all cases the result is added to the user’s gene cart as gene lists.

**Gene Network Analysis Tools: Gene Network** analysis allows users to query our Multinetwork data and displays the results in a graph using Cytoscape, an open source project that we have built upon. The output of the analysis is a subgraph of the Multinetwork containing only the genes that are in the query list. The tool also allows users to include and exclude specific interactions before displaying the graph [[1](#_ENREF_1),[2](#_ENREF_2)]. **Supernode Network** analysis can be used to help summarize the results of a Gene Network analysis. The genes in the gene network are grouped into “Supernodes” based on their functional annotation and they are associated with other Supernodes with edges determined from the Gene Network data. A gene can be present in more than one Supernode. The number of genes it contains determines the size of the Supernode. VirtualPlant also allows calculates simple **network statistics** such as gene “hubbiness”, which helps suggest candidates for future studies.

**Cross-Species Analysis Tools:** There are many methods and databases that can be used to identify homologous genes from one species to another. **Homology Mapping** allows Biologists to select the method or database they prefer to map their gene list of interest to another species. Biologists can also use the tool to make Gene Networks as described in the case study below. Alternatively, Biologists can align their sequence of interest to one of the fully sequenced genomes using **BLAST** to identify the homologs.The BLAST results can be added to the gene cart as a gene list.

**VirtualPlant Flexibility**: VirtualPlant is modular, not only because it can support different types of data and tools, but also because VirtualPlant allows different species to be added to the VirtualPlant database. Resources such as PlantGDB (http://www.plantgdb.org) and Phytozome ([www.phytozome.net](http://www.phytozome.net)) provide data for different species in a specific format. We have created functions parse this format and can automatically create a VirtualPlant database for a species hosted by Phytozome.

The VirtualPlant database contains gene annotations and many common ontology terms such as Gene Ontology and Plant Ontology. The Arabidopsis multinetwork includes metabolic pathways from KEGG and ARACYC, protein-protein interactions from BIND and Interolog databases (see Table X for a complete listing of data sources). The database also contains publicly available Microarray experiments obtained from NASC and GEO.

**Software and Data Availability**: VirtualPlant is accessible via the website www.virtualplant.org. Registered users (currently > 1,000 users) store their data sets and use many tools to analyze their genomic data. The website does not require a password and free when used for non-for-profit purposes.

***Publications: Peer reviewed journal articles, chapters and books.***

**VirtualPlant: Tool development for Plant Systems Biology**

Katari M, Nowicki S, Aceituno F, Nero D, Kelfer J, Thompson L, Cabello J, Davidson R,

Goldberg A, Shasha D, Coruzzi G, Gutierrez R (2010) "VirtualPlant: A software platform

to support Systems Biology research. ***Plant Physiol***. Feb;152(2):500-15.

Nero D, Kelfer J, Katari M, Tranchina D, Coruzzi G (2009) In Silico Evaluation of Predicted

Regulatory Interactions in Arabidopsis thaliana. ***BMC Bioinformatics***. Dec 21;10(1):435.

Poultney C, Gutierrez R, Katari M, Gifford M, Paley W, Coruzzi G and Shasha D (2007) "Sungear:Interactive visualization, exploration & functional analysis of genomic

datasets". ***Bioinformatics***, 23:259-61.

Ferro A, Giugno R, Pigola G, Pulvirenti A, Skripin D, Bader G, Shasha D, "NetMatch:

a Cytoscape Plugin for Searching Biological Networks" ***Bioinformatics***,

2007 23(7):910-912.

**Applications of VirtualPlant: Hypothesis Generation and Testing**

Krouk G, Tranchina D, Lejay L, Cruikshank A, Shasha D, Coruzzi G, and Gutierrez R (2009) “A

systems approach uncovers restrictions for signal interactions regulating genome-wide responses to nutritional cues in Arabidopsis.” ***PloS Comp Biol***. Mar;5(3):e1000326. (Highly Accessed).

Gutierrez R, Stokes T, Thum K, Xu X, Obertello M, Katari M, Tanurdzic M, Dean A, Nero D,

McClung R & Coruzzi G (2008) "Systems approach identifies an organic nitrogen-

responsive gene network that is regulated by the master clock control gene CCA1"

***Proc. Natl Acad Sci USA*** 105, 4939-4944 (Faculty of 1000 Recommended: Factor 3).

Gutierrez R, Gifford M, Poultney C, Wang R, Shasha D, Coruzzi G, Crawford N (2007) "Insights

into the genomic nitrate response using genetics and the Sungear Software System"

***Journal of Experimental Botany*** doi: 10.1093/jxb/erm079.

Gutierrez R, Lejay L, Chiaromonte F, Shasha D, Coruzzi G (2007) "Qualitative network models

and genome-wide expression data define carbon/nitrogen-responsive biomodules in

Arabidopsis" ***Genome Biology***, 8: R7 (Faculty 1000 Must Read: Factor 6).

**Plant Systems Biology: Reviews, Books and Outreach**

Ruffel S, Krouk G, Coruzzi G (2009). "A Systems View of Responses to Nutritional Cues in

Arabidopsis: Towards a Paradigm Shift for Predictive Network Modeling”.

***Plant Physiol***. Nov 25 [epub ahead of print].

Gutierrez R, Coruzzi G., Eds (2009) Book: “***Plant Systems Biology***”, Annual Plant Reviews;

Blackwell Publishing: Oxford, UK, 2009, Vol. 35. 360 pages.

Coruzzi GM, Burga A, Katari MS, and Gutierrez RA (2009) “Systems Biology: Principles and

Applications in Plant Research”. In ***Plant Systems Biology***, Annual Plant Reviews;

Blackwell Publishing: Oxford, UK, 2009, Vol. 35. Pgs 3-31. Book Chapter.

Gifford M, Gutierrez R, and Coruzzi G (2006) "Modeling the Virtual Plant: A Systems Approach

to Nitrogen-Regulatory Gene Networks". Essay 12.2 Chapter 12. Assimilation of mineral

nutrients;***In: A companion to Plant Physiology,*** Fourth Edition, Lincoln Taiz and

Eduardo Zeiger, http://4e.plantphys.net/article.php?ch=12&id=352.

Gutierrez R, Shasha D and Coruzzi G. (2005) "Systems Biology for the Virtual Plant". ***Plant***

***Physiol.*** Vol 138, pp 550-554.

**EDUCATION AND TRAINING**: The development of the Systems Biology tools and the Virtual Plant software platform has trained undergraduates, MS and PhD students in Systems Biology. Students trained include **Undergraduates**: Steve Nowicki (NYU CAS), Varuni Prabhakar (Barnard College), Rebecca Davidson (BS Computer Science); **Masters Students**: Ana F. Arroja (MS student, NYU Courant), Ranjita Iyer (MS Computer Science) Jonathan Kelfer (MS Computer Science), Lee Parnell (MS Computer Science), (Jarod Wang, MS Computer Science); **PhD Students**: Chris Poultney (PhD student, NYU Courant), Jason Reisman (PhD student, NYU Courant), Saurabh Kumar (PhD student, NYU Courant). These students have gone on to PhD programs (Prabhakar, Parnell) as well as to industry (Kelfer, Bloomberg).

-----------------------------------------------------------------------------------------------------------------

**RESEARCH PLAN:** With the advent of Next-Gen sequencing, Systems Biology approaches and “network thinking” will become increasingly important especially across species. In addition, crop species should be able to build on the “network knowledge” from model species such as Arabidopsis, to enable translational research. The VirtualPlant software platform/integration environment supports Systems Biology studies across multiple species and allows integration of knowledge across experiments, data types and perhaps most importantly across species.

**Comparison of VirtualPlant to existing Network Analysis tools**. VirtualPlant is complementary to several excellent systems for plant network analysis in the plant community including: COP [[18](#_ENREF_18" \o "Ogata, 2009 #18)]; ATTEDII [[19](#_ENREF_19" \o "Obayashi, 2011 #19)]; PlaNet [[20](#_ENREF_20" \o "Mutwil, 2011 #20)]; GeneMania: [[21](#_ENREF_21" \o "Mostafavi, 2008 #21)]. These network tools offer facilities including node-vicinity networks as we show in Table X. In addition to the novel visualization tools it provides (e.g. Sungear [[16](#_ENREF_16)], BioMaps, GeneSect, etc), VirtualPlant’s two main unique features are i) the GeneCart and ii) the ability to build a network from scratch. The GeneCart supports iterative exploration of data sets by allowing a researcher to apply a tool, return the result to the Gene Cart, then apply a different tool on the result and this can continue, supporting the iterative process of Systems Biology inquiry. The GeneCart can hold gene lists, experiments, and networks. The ability to build a network from scratch is important, because biological interactions are context-specific. Depending on the developmental stage and environmental factors, only certain interactions will be observed. For example, as we show in Section 2, starting from nitrogen-treatment experiments in Maize and Arabidopsis, we create a synthesized network in Maize that exploits the Arabidopsis multinetwork to derive refined networks and hypotheses from Maize expression data. The incorporation of dynamic transcriptomic data with static interaction data gives us a network that better models the molecular circuitry underlying the nitrogen response, as compared to only a correlation network either in Arabidopsis or Maize.. What makes VirtualPlant unique is, as we have seen above, the Gene Cart function enables iterative exploration, as well as some of our analysis tools including homology based (e.g….XXXXXX). In addition, we offer researchers the ability to build their own networks in some species of interest and while capitalizing on experiments and data in data rich species such as Arabidopsis.

**Table X**. **Comparison of network analysis platforms complementary to VirtualPlant**. Most platforms listed feature a pre-constructed network that a user can query with one or more genes. PlaNet [[20](#_ENREF_20)], GeneMania [[21](#_ENREF_21)], and VirtualPlant each host data from multiple species, but Virtual Plant is unique and complementary in that it enables network merges across species through homology mapping. VirtualPlant is centered on user-generated data and is the only current plant network tool that can be used to build a context specific network (Build Your Own Network) from a user-selected set of experiments.

**Section 1: Scientific value of Virtual Plant to enabling Systems Biology studies across the plant genomics community. *Rationale***: In this aim, we discuss the prior and anticipated impact of VirtualPlant on enabling Systems Biology research within the plant genomics community.

**1.** **VirtualPlant (VP) Users**. VirtualPlant was specifically designed to enable plant biologists with no computer expertise to analyze their own data - enabling valuable insights from a plant biologist perspective. This user base includes researchers from academe and industry. VirtualPlant keeps track of “function calls” and “user activity” to better understand which features are most useful to our user base. From these statistics we can also determine how many users are actively using VirtualPlant. As of July 2012, there are 1,033 registered VirtualPlant accounts. In Fig. X, we present VP user demographics as revealed by the registered email addresses. Among the geographically discernible domains, the largest group of users are from .edu domain, which generally represents accounts from users based in the US academia. Following the US (.com and .edu), uk (United Kingdom), fr (France), de (Germany), cl (Chile), and au (Australia) are the most frequent users. Registration is not mandatory to access VirtualPlant tools, and 44% of function calls are made by anonymous users. However only registered users can save their data in their “Gene Cart” and analyze their experiments. For this reason, registered email addresses represent the more active, recurrent VirtualPlant users. The .com domain, in addition to personal accounts of unknown users, includes accounts from companies such as Monsanto and Syngenta (data not shown to preserve privacy). Within the United States, researchers from numerous prestigious universities currently use VirtualPlant (Fig. X). In addition to the host institution, New York University, the largest numbers of VP users in academe are from University of California at Davis, Stanford University, Duke University, Iowa State University, and other first-rate universities.

**Testimonials from several of VirtualPlant high-end users include**: [Dennis took away Mary. We already use her above.] ***Sibohan Brady (UC Davis) writes****: “I work on gene regulatory networks in plants, and have published several recent papers that comprise the mapping and analysis of these networks [[5](#_ENREF_5" \o "Brady, 2011 #6),[13](#_ENREF_13" \o "Gaudinier, 2011 #17)]. Positive aspects of Virtual Plant and its tools for Arabidopsis research include the ability to work with lists of genes, to perform functional and network analysis between multiple gene lists, to normalize different .CEL files from the GEO database, and to save your lists of genes in GeneCart in your profile for later use. In the process of submitting these manuscripts, a requirement for Molecular Systems Biology was to have all protein- DNA interactions hosted in a public database for free download. None of the standard interaction databases would accept our interactions since they were in yeast one hybrid or genetic interaction format. Instead, Virtual Plant was one of two plant-based websites that accepted this data. It was very easy to work with their administrator to have our data uploaded in the multinetwork, and I am very grateful that they provide this resource in addition to many other fantastic tools that my lab uses regularly.”*

**2. VirtualPlant Usage.** VirtualPlant includes a suite of tools that are useful to analyze –omics datasets. Fig. X shows a barplot of the different tools and the frequency of their usage by a registered user. Since some of the tools are recent additions, only activity since the start of 2012 is shown. Here we see that BioMaps (gene list enrichment analysis of GO and MIPS terms) and set operations (intersection, union of genesets) are the most highly accessed features of VirtualPlant. These popular tools are followed by the number of times registered users have logged in to use “Sungear” (an interactive java application for comparing multiple gene sets) [[16](#_ENREF_16)] or “GeneSect” (statistical evaluation of the overlap in gene sets). These results demonstrate that the most common tools in genomic research today are the ability to compare gene lists, most likely derived from different conditions, experiments or literature, and enrichment analysis such as GO or MIPS terms. These simple operations comparing gene lists allow the biologist to qualitatively and quantitately measure the effect of a treatment or a gene knockout. Another observation based on tool usage is the “Network Analysis” feature in VirtualPlant is used more often compared to “Microarray Processing”, implying that VirtualPlant is very valuable as a post-microarray processing analysis tool to investigate relationships between genes and to derive hypothesis from networks formed.

**3. VirtualPlant Citations.** VirtualPlant and its tools have been cited or referenced in 50 publications since June 2006, when it was made public during the International Conference of Arabidopsis Research (ICAR). Fig. X shows a linear trend representing an increase in the number of publications that cite VirtualPlant, thus representing the growing impact VirtualPlant is having on the plant community since its release. Data for the graph was collected from Google Scholar. All publications that cite VirtualPlant [[1](#_ENREF_1)], Sungear [[16](#_ENREF_16)] or simply use “www.virtualplant.org” in their text were counted. A closer look identified which tool on VirtualPlant was used in the analysis, and this information is represented in the color codes shown in Fig. X.

**Expected Scientific Value for the Future**: With the inclusion of more species, we expect VirtualPlant to have a multiplicatively broader impact to the plant community, because translational studies will become the norm. In addition, VirtualPlant’s tools can be incorporated into other systems. For example, we recently worked with iPlant to create a stand-alone version of BioMaps to be integrated into the iPlant Discovery Environment where the tools will be available to all Discovery Environment users. Finally, VirtualPlant separates “Microarray Processing” from “Network Analysis”, enabling VP to grow and thrive as new forms of expression data evolve (e.g. RNA-Seq), and will enable it to support future changes in data types.

**Section 2: VirtualPlant: Enabling Translational Research. *Rationale:*** Agronomically important plant species have now been inducted into VirtualPlant including Rice and Maize. While integrating this data, we also created new ways to support plant biologists in their efforts to perform cross-species comparisons when analyzing their data. Such comparisons are critical to leverage the large amount of genome interaction information available in some of the model plant species, such as *Arabidopsis thaliana* and *Oryza sativa*. To demonstrate how these new databases and tools can be used to translate knowledge from one species to another, especially to leverage Arabidopsis data for translational studies in crops, we perform a case study using a publicly available microarray dataset in Maize. We selected the dataset from Yang et al. 2011 [[22](#_ENREF_22)] that discovered sensitive gene expression biomarkers for the *in planta* nitrogen status of maize.  To illustrate the cross-species functionality, we present a concise step-by-step walkthrough of a VirtualPlant analysis for this Maize dataset and its comparison to Arabidopsis.

**A Case study of Maize Translational Networks derived from VirtualPlant.**

The steps below conducted in VirtualPlant are illustrated in Fig. X

1. **The Maize Data:** The VirtualPlant database for *Zea mays* was created using two publicly available sources (Phytozome v8.0 [http://www.phytozome.net] and Maize Sequence [http://maizesequence.org]). The Maize microarray dataset used in this case study contains a total of 90 samples from nitrogen-treated plants [[22](#_ENREF_22)]. The experimental design of this study includes metadata such as time-of-day, genotype, and nitrogen-treatment, allowing for the investigation of multiple questions. For the purposes of this case study, we are interested in genes that are differentially expressed in response to nitrogen-treatment at two different time points on two days (10AM /11PM). This analysis opens a window into the interaction of nitrogen with the circadian clock, a relationship discovered in Arabidopsis using the Multinetwork [[3](#_ENREF_3)].
2. **The Maize Correlation Network:** Using the current functions in VirtualPlant Maize, the maize N-treatment microarray data was normalized using the RMA method [[23](#_ENREF_23)], followed by 2-way ANOVA. This analysis identified genes that are differentially regulated in response to either Nitrogen-treatment, or Time, or the interaction of Nitrogen-treatment and Time. Affy Probes with an FDR adjusted p-value < 0.05 were considered to be differentially expressed. Using automated functions in VP Maize, Affy probes were mapped to Maize annotated genes based on BLAST alignments of Affy probe sequences to the Maize annotated transcripts. [Dennis says…..this sounds like development: Currently VirtualPlant performs the analysis at the gene locus level, but in the future we will be able to support distinct transcripts from a single locus, as such information can be obtained from Next-gen sequences.] Next, VP Maize was used to generate pairwise gene correlation using the Pearson method, and all gene pair correlation values with a p-value < 0.05 were used to create a Maize gene expression correlation network. This maize correlation network consisted of 5,000 genes, whose expression is correlated under N-treatment conditions (Fig. X, 2). This network is both too large to enable focused hypothesis generation, and most of the maize genes in the network were unannotated (>50% hypothetical or unknown genes), making hypothesis generation impossible. To aid the interpretation of the Maize correlation network, we analyzed the Maize correlation in the context of i) the knowledge of gene interactions in the Arabidopsis multinetwork, and ii) correlation networks based on Arabidopsis N-responsive genes. This cross-species analysis enabled us to generate a specific hypothesis for conserved N-regulatory networks in Maize that included several TF hubs that had previously been validated to regulate N-assimilation in Arabidopsis, illustrating the use of translational studies to gain insight.
3. **The Arabidopsis Data:** The VirtualPlant database for the *Arabidopsis thaliana* multinetwork encompasses gene, protein and RNA interaction data from many different sources (See Table X) [[1](#_ENREF_1),[2](#_ENREF_2)]. The Arabidopsis microarray dataset we selected to compare with the Maize N-treatment dataset is from our highly accessed Arabidopsis study which identified N-regulatory networks for which key hubs that were subsequently experimentally validated [[3](#_ENREF_3)].
4. **The Arabidopsis N-regulatory Network:** The 863 N-regulated genes in Arabidopsis [[3](#_ENREF_3)] were used to create an Arabidopsis correlation network (correlation > 0.8 or -0.8, p val <0.05) (Fig. X, 4A) as in [[3](#_ENREF_3)]. The 863 N-regulated genes were also used to query the Arabidopsis MultiNetwork (Fig. X, 4B), which includes multiple edge types including TF🡪target genes based on the presence of cis-binding elements determined from AGRIS and TransFac (Table X), resulting in a N-regulatory sub-network of 828 genes. Next, we used VirtualPlant to generate an intersection of these two networks (N-regulated correlation and N-regulated multinetwork) (Fig. X, 4C) yielding a N-regulated subnetwork with 613 genes, in which genes were connected by co-expression and by at least one other interaction edge (e.g. cis-binding, protein:protein interaction) (Fig. X, 4C).
5. **Creating a Maize Translational Network from Arabidopsis**: The Arabidopsis N-regulated sub-network of 613 genes (Fig. X 4C), was next compared to the Maize genome based on a homology map function (best hit from Phytozome, www.phytozome.net) which is a tool in VirtualPlant. This homology comparison uncovered 970 genes in a homologous Maize N-network (Fig. X, 5A). An intersection of this Homology-based Maize N-network (970 genes) (Fig. X, 5A) with the Maize N-regulated correlation network (5,000 genes, Fig. X, 2) (from Step 2), uncovered a network of 42 Maize N-regulated genes in an interaction network that is conserved across Maize and Arabidopsis. The Arabidopsis network knowledge not only helped to focus the genes for translational studies, but also was crucial to network annotation, as it annotated 50% of the genes in the network which were hypothetical or unknown genes in Maize (Fig. X). These 42 N-regulated genes interactions are therefore present in both the Arabidopsis N-regulated network (Fig. X, 4C), in homology Maize network (Fig. X, 5A) and in the N-responsive Maize correlation network (Fig. X, 2). This list of 42 N-regulated and interacting genes in the “Maize Translational Network” includes 6 transcription factor hubs (CCA1, GLK1, LHY1, ARF7, GLK1, Bip9, HB16), and target genes including N-assimilation genes such as a nitrate transporter (NRT1.5) and nitrate reductase (NIA1).
6. **Validation of Maize Translational N-regulatory network:** The nature of the genes in the highly conserved N-regulatory network conserved between Maize and Arabidopsis was only possible to identify due to: i) Arabidopsis annotation, ii) Arabidopsis interactions, and iii) a comparison of correlation networks across Maize and Arabidopsis. This analysis enabled us to identify and help annotate a small network (42 genes) of N-regulated Maize genes, several of which are validated genes involved in N-regulation in Arabidopsis [[3](#_ENREF_3" \o "Gutierrez, 2008 #3)]. This fairly small list of genes can now be the focus for translational studies between Arabidopsis and Maize. This contains several transcription factor (CCA1 and LHY) hubs of an N-regulatory network conserved between corn and Arabidopsis. In a previous study, these TFs were validated to be involved the N-regulation of circadian control [[3](#_ENREF_3" \o "Gutierrez, 2008 #3)]. In addition to these regulatory genes, genes involved in nitrogen uptake and reduction were also found in this conserved N-regulatory networks, including a Nitrate Transporter 1.5 and Nitrate reductase. [dennis thinks we should say if genes of unknown function are also found]. These results show that comparative network analysis betw**e**en Arabidopsis and Maize is beneficial for translational research to focus on conserved regulatory networks, and to help identify gene function and interactions in data poor crop genomes (Maize). By interpreting the Maize expression correlation network, in the context of the Arabidopsis multinetwork and expression data, we are able to identify the most important players of nitrogen regulation. That these conserved TFs and network modules are functional in regulating genes involved in Nitrogen assimilation has already been experimentally validated [[3](#_ENREF_3" \o "Gutierrez, 2008 #3)]. Conserved networks are analogous to conserved sequences – the more distant the species, the less conservation one expects to find. However the conserved networks in these distantly related species are highly likely to have a function.

IS THIS REDUNDANT TO ABOVE? A biological network is a representation of the molecular interactions that exist in the cell. Biological networks are highly responsive to signals from within the organism and from the environment. Thus, to best model the relevant Biological network, VirtualPlant allows the Biologists to create a network based on the conditions that they are interested in. The overlap between the Maize and the Arabidopsis Nitrogen networks may be small but the interactions represent conserved Nitrogen and circadian regulatory interactions, which is one of the questions the Maize experimental dataset was designed to discover.

**Testimonials on need for Crop Species from VirtualPlant users: Mary Wildermuth (UC Berkeley) *“****With the availability of genomic resources for a variety of plants including global, high quality expression data, the addition of other species to VirtualPlant would be very welcome and is critical to its growth and utility. The ability to readily move back and forth in analyses among different species allows for increased confidence in hypotheses as well as more rapid translation of work in model organisms to agronomic species. In my COPB review (2010) [*[*6*](#_ENREF_6)*], I analyzed gene expression data associated with locally increased endoploidy for a variety of systems.  At that time, I wished VirtualPlant included these other species!* “ Sibohan Brady (UC Davis) writs: *“Our lab will soon be generating extensive tomato gene co-expression maps, and would love to be able to deposit such data in Virtual Plant as well as to query our data against other existing tomato data which has just been recently published.”*

**Expected outcomes of Aim 2:** The case study demonstrates how VirtualPlant enables Systems Biology research in crop species, and can be augmented by what is already known in model plant species. The availability of such a tool will empower not only academic researchers, but also the agricultural companies in identifying master regulator genes of plant response to stress or infection, and also genes that can be used as biomarkers to test a given crop’s status. VirtualPlant will be kept updated by including more and more species as their genomes and annotations reach maturation and as more global expression assays are developed for each of the newer species. Examples of crop genomes that will be included during this sustaining grant are: Glycine max, Medicago, Sorghum, Poplar, Brassica, Tomato. We have reached a fair degree of automation in this process by relying on standards of genome and annotation data that are evolving at data-generating centers such as JGI, PlantGDB.

**Section 3: VirtualPlant: Dissemination and User support/interaction:** The VirtualPlant platform has been disseminated to the Arabidopsis community on numerous occasions starting with its debut at the ICAR (International Conference on Arabidopsis Research) meeting in 2006 Please see below a list of presentations about VirtualPlant made by PI and Co-PIs. With the introduction of Maize and Soybean into VirtualPlant, we intend to inform and involve these new plant communities over the course of this renewal as follows:

**Workshops**: An interactive workshop to introduce VirtualPlant and publicize its features will be held at the upcoming Maize genetics conferences (2013-2015) and at the Molecular & Cellular Biology of the Soybean Conference 2014. These brief (30 minutes- 1 hour) workshops, will be led by the Co-PI Manpreet Katari and aim to provide Corn and Soybean geneticists with a hands-on experience of uploading data and analyzing test data (or their own experimental data) in the VirtualPlant environment. In addition to helping advertise the VirtualPlant platform, these sessions will give us crucial feedback on the usefulness of VirtualPlant features and allow us to make support decisions.

**Seminars/Webinars**: VirtualPlant and its tools are regularly presented at major plant biology conferences such as ICAR. In addition, open webinars followed by discussion sessions, will be held annually to better communicate with and support the users.

**Support and feedback**: Continued support is provided to the users through the help section of the VirtualPlant website and a VP users mailing list (virtualplant-discuss@googlegroups.com). The VirtualPlant site will soon feature a video tutorial to give new users an overview of the platform’s features and user interface. The help section of VirtualPlant also contains a feature request page through which our users can submit requests to add new features or modify existing ones. In the past, we have added many tools, such as the network view of BioMaps and GeneSect, in response to user needs. In its current form VirtualPlant is a powerful and robust platform, but we anticipate that small changes to existing tools or modifications to make them work for the newer species will be needed and we intend to address such issues in a timely manner.

**Impacting a broader scientific community:**  In order make our tools available to a broader systems biology community we are collaborating with developers of other software systems to allow us to share our tools. As mentioned, we have submitted our most popular tool (BioMaps) to iPlant Discovery Environment and are awaiting approval. We are also working with developers of Gaggle to integrate VirtualPlant and its tools with the Gaggle system, which will expose VirtualPlant to systems biologists who do not necessarily work on plants, thus making an impact on a broader scientific community.

**Timeline: Year 1**: Provide FREE online Webinars, Organize VirtualMaize workshop at the 2013 Maize genetics meeting. Semi-automated induction of new species into VirtualPlant via Phytozome. Launch VirtualSoy and Virtual Medicago. Update data for all supported species. **Year 2:** Provide more FREE online Webinars. Organize VirtualSoy workshop at the 15th Biennial Soybean Conference. Launch VirtualGrape and VirtualPopulus. Update data for all supported species. **Year 3**: Provide Free online Webinars, Organize a VirtualPlant workshop at 2015 Plant and Animal Genomes meeting. Update data for all supported species. Launch Virtual Sorghum, Tomato and Brassica.

**PLAN TO INTEGRATE RESEARCH AND EDUCATION**:

**Cross Training of Biologists and Computer Scientist in Systems Biology**. The development of Systems Biology tools in this project has and will involve biologists teaching computer scientists about topics like genetics, experimental genomics, and the computational challenges of analyzing genomic data. We do this informally at our weekly joint lab meetings at which graduate students and post docs from NYU Biology and NYU Courant each present their work to the group. This project involves a resident full-time systems administrator (Roberto Jimeno) working within a Biology lab, interacting closely with wet-bench biologists. The PI computer scientists (Shasha and Katari), are also involved in training and engaging computer scientist students at all levels in the emerging field of Systems Biology. In the last year, they have trained two PhD students, two interns and two MS students from Courant working in this environment. For a complete listing of students trained in the past 4.5 years, see Education and Training section in Results from Prior support.

**Workshops and Classroom Training in Genomics and Systems Biology**: We also provide formal training in the form of workshops and classes to enable Systems Biology. Examples of this include a once-a-year “R-Boot camp”, which involves a weekly software workshop in “R”, which aims to teach biologists how to analyze their own genomic data. A workshop on Virtual Plant has been taught two times, once by Jonathan Kelfer, a MS student working on the project and most recently by Manpreet Katari, co-PI. Students have included several faculty on sabbatical at NYU including most recently: Mary Lou Guerinot and Rob McClung of Dartmouth. Students will be exposed to Genomics and Systems Biology also through a series of formal courses offered by faculty at NYU’s Center for Genomics and Systems Biology including: BIOL-GA-1128 Systems Biology; BIOL-GA-1130 Applied Genomics: Introduction to Bioinformatics & Network Modeling; BIOL-GA-1127 Bioinformatics & Genomes; BIOL-GA-2030 Statistics in Biology; BIOL-GA-1007 Bioinformatics for Biologists; BIOL-GA-1009 Biological databases and datamining; BIOL-UA-103 Bioinformatics in Medicine and Biology; BIOL-UA-38 – Introduction to Genomics and Bioinformatics. PhD students have and will continue to present their work in the weekly PhD seminar series hosted by the Biology Department. Computational students will be involved in constructing the pipeline and making it perform through the use of parallelization. Such students will also help to develop and test optimization and machine learning algorithms for network inference.

**PLAN TO INTEGRATE DIVERSITY**: We are committed to training scientists at the graduate and postdoctoral levels who can do independent research that cuts across fields and expertise in genomics.  Our research team is also committed to diversity.  Researchers in our current and previous NSF grants included Hispanic and African-American scientists.  We will continue to actively seek out and recruit scientists from under-represented minorities to participate in our research in our continuing commitment to increase diversity in our research program. Five female scientists are associated with this project: Coruzzi (co-PI); Rebecca Davidson (Programmer); Varuni Prabhakar (UG Programmer); Ana Arroja (MS); Ranjita Iyer (MS Courant) Stuti Srivastava (MS Biology). Damion Nero a minority, recently graduated PhD student, has written programs contributing to the Virtual Plant project. Roberto Jimeno (Systems Admin) associated with this project is of Hispanic origin.

**SHARING OF RESULTS**:

**PUBLICATIONS:** As we generate VirtualPlant for the various crop species listed, we intend to publish and publicize the crop species VirtualPlant as papers in which VP is used to generate translational networks that derive hypotheses for testing. We will work on a VP Maize paper with Stephen Moose’s Maize data and on a VP Soy paper with Gustavo Macintosh.

**MANAGEMENT PLAN**: PI Gloria Coruzzi (NYU Biology) and co-PI Dennis Shasha (NYU Courant) will each supervise personnel, organization, intellectual developments and contributions. The PI Coruzzi will insure that all the project goals are met. Co-PI Shasha will serve as a computational advisor and mentor. Co-PI Manpreet Katari will serve as Project Manager to oversee the daily operations and programmer activities to ensure that the needs and concerns of the participants are addressed. As detailed below, Kranthi Varala will focus on outreach to the crop communities including regular meetings with the beta-testers described below. Virtual meetings will be held twice a semester, along with one on-site meeting per year, with our collaborator and founding member of VirtualPlant, Rodrigo Gutierrez (Chile), to evaluate work status and long-term planning.

**Project Manager: Dr.** **Manpreet Katari** (NYU Biology) will be in charge of the bioinformatics data and software aspects of the project. To enable efficient information exchange of raw and processed data, a file server has been set up at NYU to store and distribute data and its analysis among users at NYU Biology and NYU Courant. This will be maintained by **Roberto Jimeno**, the Systems Administrator for this project, who will also maintain the web server, database server, and update the multinetwork databases. **Kranthi Varala**, a Plant Biologist and Bioinformatician with experience in crop genomics, will be liaison to various crop plant communities including Maize and Soy.

**Roles of Participants:**

|  |  |  |
| --- | --- | --- |
| **Name** | **Institution** | **Role** |
| ***Gloria Coruzzi***-PI | NYU Biology | Project Leader |
| ***Dennis Shasha***-Co-PI | NYU Courant | Co-leader: Computational |
| ***Manpreet Katari***-Co-PI | NYU Biology | Project Manager |
| ***Kranthi Varala-*** *Senior Personnel* | NYU Biology | Bioinformatician & Crop Liaison |

**COORDINATION WITH OUTSIDE GROUPS:Please see attached letters of collaborations:**

**Coordination with International Colleagues**: **Rodrigo Gutierrez** (U. Catolica, Chile), the creator of the Arabidopsis multinetwork [[1](#_ENREF_1),[2](#_ENREF_2)] will assist in the assembly and updating of multinetworks for crop species including Vitis (Grape), Medicago truncatula, and Zea mays. His work will be evaluated during regular virtual meetings as well as at the annual grant progress meeting.

**Coordination with Crop Databases**: We will coordinate VirtualPlant with databases like Maizesequence.org (lead by **Doreen Ware**, CSHL) and Soybase (lead by **David Grant**, Iowa State). We will provide access to VirtualPlant tools from Gramene, Soybase etc. Further, we will attend the annual meetings and conduct VP workshops to support and assess the needs of the crop user communities. Finally, video tutorials will be hosted on the VirtualPlant site to introduce new users to the interface and tools available. We will also continue to support our existing user base through the help section on our website and a mailing list where users can ask questions.

**Coordination with Crop Species Beta Testers**: Kranthi Varala will coordinate outreach to the crop community. We have also enlisted beta testers of VirtualSoy (**Gustavo Macintosh**, Iowa State) and Virtual Maize (**Stephen Moose**, U Illinois), as per their letters of collaboration, with whom we will conduct regular meetings to assess the VirtualPlant functions and needs for these crop applications.

**Letters of Support:** We have included letters of support for VirtualPlant from two high-profile and hi-end users of Virtual Plant, **Mary Wildermuth** (UC Berkeley) and **Siobhan Brady** (UC Davis), who assess the various tools and functions of VirtualPlant as an integrated platform for Plant Systems Biology, and how it has enabled their research discoveries.

**References Cited**

1. Katari, MS, Nowicki, SD, Aceituno, FF, Nero, D, Kelfer, J, Thompson, LP, Cabello, JM, Davidson, RS, Goldberg, AP, Shasha, DE, Coruzzi, GM, and Gutierrez, RA, *VirtualPlant: a software platform to support systems biology research.* Plant Physiol, 2010. **152**(2): p. 500-515.

2. Gutierrez, RA, Lejay, LV, Dean, A, Chiaromonte, F, Shasha, DE, and Coruzzi, GM, *Qualitative network models and genome-wide expression data define carbon/nitrogen-responsive molecular machines in Arabidopsis.* Genome Biol, 2007. **8**(1): p. R7.

3. Gutierrez, RA, Stokes, TL, Thum, K, Xu, X, Obertello, M, Katari, MS, Tanurdzic, M, Dean, A, Nero, DC, McClung, CR, and Coruzzi, GM, *Systems approach identifies an organic nitrogen-responsive gene network that is regulated by the master clock control gene CCA1.* Proc Natl Acad Sci (USA), 2008. **105**(12): p. 4939-4944.

4. Gifford, ML, Dean, A, Gutierrez, RA, Coruzzi, GM, and Birnbaum, KD, *Cell-specific nitrogen responses mediate developmental plasticity.* Proc Natl Acad Sci (USA), 2008. **105**(2): p. 803-808.

5. Brady, SM, Zhang, L, Megraw, M, Martinez, NJ, Jiang, E, Yi, CS, Liu, W, Zeng, A, Taylor-Teeples, M, Kim, D, Ahnert, S, Ohler, U, Ware, D, Walhout, AJ, and Benfey, PN, *A stele-enriched gene regulatory network in the Arabidopsis root.* Mol Syst Biol, 2011. **7**: p. 459.

6. Wildermuth, MC, *Modulation of host nuclear ploidy: a common plant biotroph mechanism.* Curr Opin Plant Biol, 2010. **13**(4): p. 449-458.

7. Ford, KA, Casida, JE, Chandran, D, Gulevich, AG, Okrent, RA, Durkin, KA, Sarpong, R, Bunnelle, EM, and Wildermuth, MC, *Neonicotinoid insecticides induce salicylate-associated plant defense responses.* Proc Natl Acad Sci U S A, 2010. **107**(41): p. 17527-17532.

8. Chandran, D, Inada, N, Hather, G, Kleindt, CK, and Wildermuth, MC, *Laser microdissection of Arabidopsis cells at the powdery mildew infection site reveals site-specific processes and regulators.* Proc Natl Acad Sci U S A, 2010. **107**(1): p. 460-465.

9. Chandran, D, Tai, YC, Hather, G, Dewdney, J, Denoux, C, Burgess, DG, Ausubel, FM, Speed, TP, and Wildermuth, MC, *Temporal global expression data reveal known and novel salicylate-impacted processes and regulators mediating powdery mildew growth and reproduction on Arabidopsis.* Plant physiology, 2009. **149**(3): p. 1435-1451.

10. Popescu, SC, Popescu, GV, Bachan, S, Zhang, Z, Gerstein, M, Snyder, M, and Dinesh-Kumar, SP, *MAPK target networks in Arabidopsis thaliana revealed using functional protein microarrays.* Genes Dev, 2009. **23**(1): p. 80-92.

11. Consortium, AIM, *Evidence for Network Evolution in an Arabidopsis Interactome Map.* Science, 2011. **333**(6042): p. 601-607.

12. Ruffel, S, Krouk, G, Ristova, D, Shasha, D, Birnbaum, KD, and Coruzzi, GM, *Nitrogen economics of root foraging: transitive closure of the nitrate-cytokinin relay and distinct systemic signaling for N supply vs. demand.* Proc Natl Acad Sci U S A, 2011. **108**(45): p. 18524-18529.

13. Gaudinier, A, Zhang, L, Reece-Hoyes, JS, Taylor-Teeples, M, Pu, L, Liu, Z, Breton, G, Pruneda-Paz, JL, Kim, D, Kay, SA, Walhout, AJ, Ware, D, and Brady, SM, *Enhanced Y1H assays for Arabidopsis.* Nat Methods, 2011. **8**(12): p. 1053-1055.

14. Ideker, T, Thorsson, V, Ranish, JA, Christmas, R, Buhler, J, Eng, JK, Bumgarner, R, Goodlett, DR, Aebersold, R, and Hood, L, *Integrated genomic and proteomic analyses of a systematically perturbed metabolic network.* Science, 2001. **292**(5518): p. 929-934.

15. Gutierrez, RA, Shasha, DE, and Coruzzi, GM, *Systems biology for the virtual plant.* Plant Physiol, 2005. **138**(2): p. 550-554.

16. Poultney, CS, Gutierrez, RA, Katari, MS, Gifford, ML, Paley, WB, Coruzzi, GM, and Shasha, DE, *Sungear: interactive visualization and functional analysis of genomic datasets.* Bioinformatics, 2007. **23**(2): p. 259-261.

17. Gutierrez, RA, Gifford, ML, Poultney, C, Wang, R, Shasha, DE, Coruzzi, GM, and Crawford, NM, *Insights into the genomic nitrate response using genetics and the Sungear Software System.* J Exp Bot, 2007. **58**(9): p. 2359-2367.

18. Ogata, Y, Sakurai, N, Suzuki, H, Aoki, K, Saito, K, and Shibata, D, *The prediction of local modular structures in a co-expression network based on gene expression datasets.* Genome Inform, 2009. **23**(1): p. 117-127.

19. Obayashi, T, Nishida, K, Kasahara, K, and Kinoshita, K, *ATTED-II updates: condition-specific gene coexpression to extend coexpression analyses and applications to a broad range of flowering plants.* Plant Cell Physiol, 2011. **52**(2): p. 213-219.

20. Mutwil, M, Klie, S, Tohge, T, Giorgi, FM, Wilkins, O, Campbell, MM, Fernie, AR, Usadel, B, Nikoloski, Z, and Persson, S, *PlaNet: combined sequence and expression comparisons across plant networks derived from seven species.* Plant Cell, 2011. **23**(3): p. 895-910.

21. Mostafavi, S, Ray, D, Warde-Farley, D, Grouios, C, and Morris, Q, *GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function.* Genome Biology, 2008. **9 Suppl 1**: p. S4.

22. Yang, XS, Wu, J, Ziegler, TE, Yang, X, Zayed, A, Rajani, MS, Zhou, D, Basra, AS, Schachtman, DP, Peng, M, Armstrong, CL, Caldo, RA, Morrell, JA, Lacy, M, and Staub, JM, *Gene expression biomarkers provide sensitive indicators of in planta nitrogen status in maize.* Plant physiology, 2011. **157**(4): p. 1841-1852.

23. Wu, Z, Irizarry, RA, Gentleman, R, Martinez-Murillo, F, and Spencer, F, *A Model-Based Background Adjustment for Oligonucleotide Expression Arrays.* Journal of the American Statistical Association, 2004. **99**(468): p. 909-917.