**PROJECT SUMMARY** **NSF Plant Genome: CrossPlant:** Cross-Species Network inference, Gene Identification, and Data Pipelines to enhance translational gene discovery.

1. **Senior personnel PI:** Gloria Coruzzi (NYU Biology, Center for Genomics & Systems Biology)

**Co-PI:** Dennis Shasha (NYU Courant Institute of Mathematical Sciences)

 **Co-PI:** Manpreet Katari (NYU Biology, Center for Genomics & Systems Biology)

 **Collaborator:** Rodrigo Gutierrez (Pontificia Universidad Catolica de Chile)

**2. Intellectual merit of the proposed activity**: Biologists in this Next-Gen era face a challenge: sequencing is relatively easy, but doing experiments on newly sequenced species is costly in time and money. For this reason, it would be useful (i) to infer networks on poorly studied species based on well-studied species and (ii) to perform mutagenesis experiments on model species based on candidates derived from the other species (e.g. crop species) of interest. Both goals would enhance translational discoveries from models to crops. Because these points (i) and (ii) are only some possible uses of networks on multiple species, another useful goal is to construct single or multi-species multinetworks by inference and/or b exploiting all the plant genome data that is currently available. ***To achieve these goals, we divide the work into three aims***

**Aim 1: Development of InferNet: A machine learning approach to use networks in data-rich species to infer networks in data-poor species.** We will develop a machine-learning approach called ***InferNET*** to *infer*regulatory networks in a data-poor target species (whether crop or non-crop), based on learning a set of rules for inferring *gene networks* based on several data-rich species including crops and/or models. These studies should enhance positive predictions of targets for translational studies between models and crops.

**Aim 2: Development of a trait-network-gene discovery pipeline:** We propose to utilize gene networks, built on conserved co-expression data from multiple crop species, to discover genes central to a particular trait of interest. This pipeline entails the construction of an "elected" network based on crops to suggest genes and gene modules to manipulate in Arabidopsis either through over-expression or knock-ins. This approach in essence mines genomic data from crop species to inform hypothesis derivation and validation in the model Arabidopsis.

**Aim 3: Development of “X-Net”**: “X-Net” enables users to construct either experimentally derived networks or inferred networks to derive both (i) species-specific networks consisting of multiple edge types (multinetworks, for short) and (ii) cross-species “elected” networks (“elected networks”, for short). Such networks synthesize knowledge within and across species to identify “cross-species” vs. species- or clade-specific network modules for hypothesis testing.

**2. Justification for NSF Plant Genome Goals**:

**(1) Development of tools and resources for plant genome research including novel technologies and analysis tools that will enable discovery**. Aim 1 will develop approaches, tools and pipelines to perform network analysis on any species or combination of species, and to exploit the large amount of data on well studied plant species to infer networks on new and emerging species. This work will achieve one of the main goals of Systems Biology – predicting network states under untested conditions.

(**2) Genomics-empowered plant research to tackle fundamental questions in plant and agricultural sciences on a genome-wide scale**. In Aim 2, we will identify network modules associated with traits in crops and test genes in those modules in Arabidopsis. In 2A, we develop and validate the approach using a mutant-rich dataset (seed development), and in 2B, apply the approach to the trait of Nitrogen use, exploiting data across a range of crop species.

This project will leverage the facilities of the current VirtualPlant software platform ([www.virtualplant.org](http://www.virtualplant.org)) developed under an NSF Grant (DBI-0445666), that includes Arabidopsis multinetwork data, analysis, integration and manipulation tools.

**3. Broader impacts of the proposed research**

**Novel training opportunities in plant genomics and bioinformatics:** 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. In addition to scientific results, this collaboration extends to joint training of biologists and computer scientists in the field of Systems Biology. The Systems Biology tools resulting from this project will empower plant biologists to use genomic data to predict a spectrum of gene networks in biology with broad applications to agriculture, the environment, and health.

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

**NSF Plant Genome (GOALS)**

This program is a continuation of the Plant Genome Research Program (PGRP) that began in FY 1998 as part of the National Plant Genome Initiative (NPGI).  Since the inception of the NPGI and the PGRP, there has been a tremendous increase in the availability of functional genomics tools and sequence resources for use in the study of key crop plants and their models.

Proposals are welcomed that build on these resources to develop conceptually new and different ideas and strategies to address grand challenge questions in plants of economic importance on a genome-wide scale.  There is also a continued need for the development of novel and creative tools to facilitate new experimental approaches or new ways of analyzing genomic data.  Especially encouraged are proposals that provide strong and novel training opportunities integral to the research plan and **particularly across disciplines that include, but are not limited to, plant physiology, quantitative genetics, biochemistry, bioinformatics and engineering.**

Three kinds of activity will be supported in FY 2012:

(1) Genomics-empowered plant research to tackle fundamental questions in plant and agricultural sciences on a genome-wide scale;

(2) Development of tools and resources for plant genome research including novel technologies and analysis tools that will enable discovery; and

(3) Mid-Career Investigator Awards in Plant Genome Research (MCA-PGR) to increase participation of investigators trained primarily in fields other than plant genomics.

**Proposals addressing these opportunities are welcomed at all scales, from single-investigator projects through multi-investigator, multi-institution projects, commensurate with the scope of the work proposed.**