**PROJECT SUMMARY** **NSF Plant Genome: “X-Net:** Integrating -omic data across-species to infer networks and enhance model-to-crop gene discovery”

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

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

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

**Senior Personnel**: Rob Martienssen (CSHL) ; Stephen Moose (U. Illinois).

**2. Intellectual merit of the proposed activity**: Plant biologists in the Next-Gen sequencing era face a challenge: sequencing is relatively easy, but doing experiments on newly sequenced or emerging crop 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 candidate genes inferred from multi-species networks. Because there could be many possible uses of networks on multiple species, it would also be useful (iii) to provide a tool to construct single or multi-species multi-networks by inference on-the fly by 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*.** *InferNET* is a machine-learning approach that will *infer*regulatory networks in a *data-poor* target species (whether crop or non-crop), based on learning a set of rules from several *data-rich* species, including crops and/or models. This *learning* approach can be used to mine data-rich species to enhance positive predictions across crops and other species.

**Aim 2: Development of a trait-network-gene discovery pipeline.** We propose to create “weighted” gene networks, built on conserved co-expression data from multiple crop species, to inform discoveries of genes central to a particular trait of interest. Target genes will be manipulated either through over-expression or knock-ins, initially using a medium through-put transient assay system in Arabidopsis, as a pipeline to whole plants. To validate the “translatability” of this approach, we will we will take the candidates that worked best in Arabidopsis and mutagenesize them on maize (Aim 2B).

**Aim 3: Development of “X-Net”**. The “X-Net” platform will enable users to construct either experimentally derived or inferred (i) species-specific networks consisting of multiple edge types and (ii) cross-species “weighted” networks for any species of interest. This platform will also automate pipelines for network approachs in Aims 1 and 2. These approaches synthesize knowledge within and across species to identify 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**. We will develop approaches, tools, and pipelines to perform network analysis on any species or combination of species (hence the name X-Net), and to exploit the large amount of data on well-studied plant species to infer networks on new and emerging species. This work will enhance translational research and 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 gene network modules associated with traits using data from crops to create weighted networks and identify candidate genes. We will develop and validate the approach using a mutant-rich dataset (seed development) amenable to testing in first model species and, then, selectively in crop species to validate translatability.

**3. Broader impacts of the proposed research: Enhance translation from models to Crops:** This project will provide a platform to the plant community that will enable plant biologists to mine the vast wealth of data accumulated across plant genomes (i) to infer networks in new and emerging crop species, and (ii) to mine crop data to inform functional studies in Arabidopsis that can then be applied to crops.  **Novel training opportunities in Systems Biology:** This project is the result of a long-standing and highly successful collaboration between biologists and computer scientists at NYU's Courant Institute. 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 biologists to use genomic data to predict and validate a spectrum of gene networks in biology –beyond plants - with broad applications to agriculture, the environment, and health.