**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)

**Co-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); Kranthi ....

**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 identify candidate genes inferred from multi-species networks for experimentation. Because there could be many possible uses of networks on multiple species, it would also be useful (iii) to provide a platform 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*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 machine learning approach can be used to mine data-rich species to perform predictions on 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 initially using a medium through-put transient assay system called “*Network Walking*” in Arabidopsis, then in whole plant studies (e.g. knock-outs and knock-ins). Finally, we will prioritize genes from Arabidopsis tests for validation in Maize as proof-of-translatability using the traits “Seed Development” and “Nitrogen-Use” as case studies of agronomic interest.

**Aim 3: Development of “X-Net Builder”**. The “X-Net” builder 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 set of species of interest, using approaches from Aims 1 & 2. X-Net will enable researchers to synthesize knowledge within and across species of interest to identify network modules for hypothesis derivation and testing.

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

**1. Development of novel 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 first in model species and, then, selectively in crop species as proof-of-principle to validate translatability.

**3. Enhance translation from models to Crops:** The X-Net platform 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, (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 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- at all levels including High School- in Systems Biology. The tools resulting from this project will empower biologists to mine the large and rapidly growing amount of existing genome data to predict and validate a spectrum of gene networks in biology – beyond plants – with broad applications to agriculture, the environment, and health.