**PROJECT SUMMARY** **NSF Plant Genome: “X-NET”:** A platform to mine cross-species network knowledge to infer and create networks that enhance translational gene discovery in crops.

**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)

**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 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 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 multinetworks by inference on-the fly and/or 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: A machine learning approach to use networks in data-rich species to infer networks in data-poor species.** ***InferNET*** 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 will enhance positive predictions of targets in 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. Those genes will then be manipulated 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 the “X-Net” Platform**: “X-Net” platform will enable 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**. This grant 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 network modules associated with traits in crops and test genes in those modules in Arabidopsis. We develop and validate the approach using a mutant-rich dataset (seed development), and validate the activity of the candidate genes.

**3. Broader impacts of the proposed research: Enhance translation from models to Crops:** This project will provide the plant community with a platform that will enable plant biologists to mine the vast wealth of data accumulated across plant genomes to (i) infer networks in new and emerging crop species, (ii) to mine crop data to inform functional studies in Arabidopsis that can then apply to plants.  **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 plant biologists to use genomic data to predict a spectrum of gene networks in biology –beyond plants- with broad applications to agriculture, the environment, and health.