**“XG-Boosting” of biofuel production in low N-by-W environments**

Gloria Coruzzi, New York University (Principal Investigator)

Dennis Shasha, Courant Institute of Mathematical Sciences – NYU (Co-Investigator)

Jean-Michel Ane, University of Wisconsin - Madison (Co-Investigator)

Nitrogen (N) and Water (W) - two essential plant resources for plant growth - are increasingly limited in marginal soils world-wide. In an effort to develop biofuel crops that can thrive in a changing environment, we will identify key genes and regulatory networks that respond to N-by-W interactions to control crop biomass, W- and N-Use efficiency (NUE and WUE). While agronomists have long-known the synergistic effect of N-by-W on crop outcomes, the molecular basis for this remains unknown. This is because traditional experimental designs uncover either W- or N-response genes, not their interactions. We address this knowledge-gap by exploiting a novel N-by-W matrix which can distinguish plant responses to; N-moles, W-volume, N/W (molarity) or synergistic NxW interactions. We will use this N-by-W response dataset - gene expression and phenotypes, captured across genotypes of varying drought sensitivity - to drive machine learning to identify *causal relationships* between transcription factors🡪target gene(s)🡪 NUE, WUE and biomass phenotypes. We will do this in the model monocot, *Brachypodium* to inform and prioritize validation studies in the biofuel crop *Sorghum*.

 **Aim 1. Quantify N-by-W responses over *TIME* in *Brachypodium* and biofuel crop *Sorghum*.** We will quantify N-by-W responses by exposing *Brachypodium* and *Sorghum* to a 4-by-4 N-by-W matrix that comprises conditions of increasing nitrogen deprivation and drought. We will collect gene expression and phenotype data (NUE, WUE, and biomass) in time-series. This time-series data will drive our machine learning method called OutPredict, to predict the causal relationship between transcription factors and target genes in Aim 3A. We will also use linear models to identify genes and phenotypes that respond to: N-moles, W-volume, N/W (molarity) or NxW (synergy) for use in Aim 3B.

 **Aim 2. Quantify N-by-W responses across plant genotypes with varying drought sensitivity.** To capture a range of NUE, WUE, and biomass phenotypes, we will test the N-by-W responses (expression and phenotype) across 9 *Brachypodium* and 3 *Sorghum* varieties that vary in their drought resistance using a N-by-W matrix in a 2-by-2 design of N-moles and W-volume (e.g. Low vs. High). The genotype data will be used to quantify the effect of genotype on NUE, WUE, and biomass. As the N and W doses in this matrix are binary (e.g. High vs. Low), we will use ANOVA analysis to identify responses to N, W or N-by-W interactions. This data will be used to drive an ensemble of machine learning methods to derive genes of “importance” to phenotypes of NUE, WUE and biomass in each species in Aim 3B.

 **Aim 3.** **Apply machine learning methods to identify causality of N-by-W response genes to NUE, WUE and biomass phenotypes**. The datasets from Aims 1&2 will drive machine learning methods to predict (1) causal relationships between transcription factors and target genes based on time-series data (Aim 3A); (2) genes that influence phenotypes based on genotype data (Aim 3B). These combined analyses will result in a ranked list of transcription factor🡪target gene(s)🡪phenotype relationships. Cross-species transfer learning between *Brachypodium* and *Sorghum* (Aim 3C) will prioritize validation studies.

 **Aim 4: Functional validation of candidate genes for adaptation of biofuel crops to low-N/low-W soils.** The top-ranked genes from Aim 3 will be validated *in planta* in *Brachypodium* and *Sorghum* using loss-of-function and overexpression mutants tested in a 2-by-2 N-by-W matrix for phenotyping. Candidate genes in *Brachypodium*, that result in significant increase in NUE, WUE, or biomass relative to wild type, will be prioritized for introduction into *Sorghum* for increasing biomass accumulation in marginal soils.

 **Outcome:** This project will provide genes that, when modified in biofuel crops, will significantly improve their biomass, WUE and NUE in arid and N-poor marginal soils. The systems biology approaches can be applied across any problem in biology and offers unique experimental design features including (1) examining the interaction between multiple environmental inputs (N-by-W) over time, (2) transfer knowledge between species: model (*Brachypodium*) to crop (*Sorghum*), and (3) new machine learning algorithms to derive causal gene regulatory networks. The PIs have complementary expertise and a history of successful collaboration and will train the next generation of students and post-docs in using genomics, systems biology and machine learning to improve crop outcome.