

## **QPSI: Quantitative Plant Science Initiative**

**Crysten E. Blaby-Haas**\*<sup>1</sup> ([cblaby@bnl.gov](mailto:cblaby@bnl.gov)), Agnidipta Ghosh<sup>2</sup>, Miriam Pasquini<sup>1</sup>, Nam Nguyen<sup>3</sup>, Sam Seaver<sup>4</sup>, Ping Zhu<sup>1</sup>, Daifeng Wang<sup>3</sup>, Steven C. Almo<sup>2</sup>, Doreen Ware<sup>5,6</sup>, Qun Liu<sup>1</sup>, Ian K. Blaby<sup>1</sup>

<sup>1</sup>Brookhaven National Laboratory, Upton, NY; <sup>2</sup>Albert Einstein College of Medicine, Bronx, NY; <sup>3</sup>Stony Brook University, Stony Brook, NY; <sup>4</sup>Argonne National Laboratory, Argonne, IL; <sup>5</sup>Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; <sup>6</sup>US Department of Agriculture, Agricultural Research Service, Ithaca, NY, USA.

**Project Goals: QPSI (Quantitative Plant Science Initiative) is a DOE-mission driven, interdisciplinary, team-based capability that aims to accelerate the acquisition of a core knowledgebase of experimentally validated protein family function. Knowing what and how to engineer and predicting the consequence of that redesign on complex systems, such as a bioenergy crop, requires a systems biology approach grounded in molecular-level understanding. Because of the lack of sequence-based function understanding that affects all BER-relevant plant genomes, our goals are to: (1) reduce uncertainty in plant protein function annotation through integrated computational and experimental approaches, (2) identify genome-based principles underlying highly conserved processes over plant evolution, and (3) define fundamental rules of sequence-function relationships that translate across bioenergy crops. The resulting knowledge will underpin genome-based functional genomics experiments, while accelerating a sequence-based understanding of genotype-to-phenotype in specific plants.**

Bioenergy crops that thrive in marginal soils and maintain performance in diverse and fluctuating environments are an essential component of a sustainable bioeconomy. Understanding and predicting biosystem productivity in these environments is hindered by a lack of sequence-to-function information. This knowledge gap impedes accomplishing Biological and Environmental Research (BER) mission to enable the design and reengineering of plants underpinning energy independence. In recognition of the critical role that knowledge of sequence-based protein function plays in the mission of BER and BSSD, Brookhaven National Laboratory (BNL) has established the Quantitative Plant Science Initiative (QPSI). The overarching goal of QPSI is to integrate experimentation with computational approaches to address the knowledge gap that exists between plant genomes and the function encoded within.

Genomics has accelerated our ability to gather systems-wide information regarding plant responses and adaptations to the environment. QPSI aims to capitalize on this cascade of genome-based data by combining scalable functional genomics and computational approaches with protein function characterization. The described activities leverage *in silico* 'omics-derived functional extrapolations and high-throughput *in vitro/in vivo* experimentation. Computational and experimental support is integrated with our *in silico* pipeline for evidence-based propagation of

annotations across bioenergy crops. The resulting gain of foundational knowledge will provide the guiding biological principles for the discovery of niche-specific and lineage-wide processes as targets for the improvement of plant-based bioenergy.

Presently, our priorities are divided into two main objectives:

Objective 1: generate evidence-based function predictions for conserved bioenergy crop protein families and provide a framework for disseminating process-level understanding of plant biology. This objective involves phylogenomic analyses of plant protein families and identification of gene regulatory networks. We are extending a computational platform, already integrated with DOE's Systems Biology Knowledgebase (KBase), for plant subsystem construction and interpretation. Subsystems provide a novel framework for *in silico* hypothesis building and testing in plants, as well as contextualized transfer of knowledge from one genome to another.

Objective 2: experimentally characterize the function of conserved plant protein families. We are generating experimental evidence for the function of conserved protein families using high-throughput (HTP) techniques. QPSI has built a custom HTP automation platform that leverages microbial phototrophs to accelerate the generation of protein family functional information. HTP protein expression/purification approaches are being leveraged to accelerate the generation of *in vitro* experimentally supported protein function evidence. We are also testing the ability of our family-centric computational and experimental functional studies to predict the phenotype of putative loss-of-function alleles in the bioenergy crop *Sorghum bicolor*. Finally, we are performing detailed characterization to provide the missing sequence-to-structure-to-function understanding for specific bioengineering targets.

Here, we present an update on the development of our capability and present progress towards our short-term objectives.

*This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER).*