

# Genomics-Enabled Plant Biology for Determination of Gene Function

Summary of projects awarded in 2019 under Funding Opportunity Announcement DE-FOA-0002060

Genomic Science Program

[genomicscience.energy.gov](http://genomicscience.energy.gov)

## Overview

The U.S. Department of Energy's (DOE) Genomic Science program, managed within the Office of Science's Office of Biological and Environmental Research (BER) supports genomics-based research and technological innovation that will lead to transformative approaches to determine and validate gene function in systems biology research on plant genomes, gene systems, and molecular processes within the DOE BER mission-space of research in bioenergy and the environment.

Developments in omics technologies have enabled unprecedented views of plants across scales, from the molecular to the ecosystem level. When integrated with computational modeling and data analysis, these technologies have the potential to greatly enhance understanding of critical plant processes such as metabolism, development, and inter/intraspecies signaling and communications. By elucidating the properties of individual components within their molecular, cellular, and organismal contexts, we hope to gain a better understanding of the living systems and to predict its behavior. Omics technologies have resulted in the generation of immense amounts of data that can be harnessed towards this goal. However, a substantial amount of computational systems biology analyses lack experimental validation of gene function and are purely correlative or computationally inferred. Current limitations in our understanding of the individual functional components of a given biological system represent a major obstacle to answering fundamental questions in plant biology. They also impede development of computational models that accurately predict the behavior of such systems in natural environments.

BER solicited applications that address the challenges and opportunities in associating gene(s) to function (i.e., genotype to phenotype) in plant systems of relevance to the BER mission in energy and the environment, with the goal of elucidating and validating the functional roles of genes, gene families, and associated pathways.

Examples of physiological and metabolic processes of interest included nutrient and water use efficiency; tolerance or resistance to abiotic stresses such as drought and

temperature extremes; developmental processes critical to enhanced biomass yield and optimization or extension of growth range; and metabolism of oils and fatty acids for biofuels and bioproducts.

Species of interest included candidate bioenergy plants (e.g., sorghum, switchgrass, *Populus*), oilseed crops (e.g., *Camelina*), as well as duckweed as a source of biofuel and bioproducts. Additionally, research using model species with relevance to and potential for translation to bioenergy crops was accepted.

Specific topics of interest included:

- Innovative approaches toward understanding the distinct functions of genes, proteins, and metabolites of the multiple cell types within complex plant organs;
- Comprehensive investigations of regulatory elements (e.g., transcription factors, cis-regulatory regions, enhancers, chromatin state, alternative splicing mechanisms) to facilitate validation of regulation of gene expression;
- Methods to discern regulatory differences among members of a gene family, particularly large homologous gene families and subsets within, and to identify gene copies likely to be functionally redundant;
- Integration of high-resolution molecular structures of proteins/enzymes and assemblies with computational modeling to identify and prioritize structural features affecting specificity;
- New comparative genomics approaches to identify reduced sets of candidate genes for detailed functional characterization across related species;
- Dissection of complex phenotypes (e.g., yield) into component genetic traits, and utilization of natural genetic variants, mutants, and transgenes for such components to characterize and validate function, with an aim towards understanding impact on the final output phenotype;
- Development and validation of methods to map quantitative associations within clades using synteny, homology, cross-species co-expression network, and other analyses.



U.S. DEPARTMENT OF  
**ENERGY**

Office of  
Science

Office of Biological and Environmental Research

## 2019 Awards

### Functional Analysis of Candidate Genes Involved in Oil Storage and Stability in Pennycress

- **PI:** Ana Paula Alonso (University of North Texas)
- **Co-Investigators:** Kent Chapman (University of North Texas) and Erich Grotewold (Michigan State University)

The goal of this research is to advance knowledge of gene function in dedicated bioenergy oilseed crops used for specialty fuels such as pennycress (*Thlaspi arvense* L.), which has promise as a source of aviation fuel. The function of candidate genes involved in lipid storage and stability will be investigated using transient expression in *Nicotiana benthamiana*, followed by validation and characterization in pennycress. Knowledge and findings generated through this project will facilitate oil production and storage in alternative oilseed crops as well as in vegetative organs of bioenergy plants.

### Coupling Metabolic Source Isotopic Pair Labeling and Genome-Wide Association for Metabolite and Gene Annotation in Plants

- **PI:** Clint Chapple (Purdue University)
- **Co-Investigator:** Brian Dilkes (Purdue University)

This project will use a robust, high-throughput approach comprising isotopic labeling, mass spectrometry, and Genome Wide Association (GWA). Coupling GWA and parallel measurements of metabolites isotopically tagged by pathways of origin will uncover novel metabolism, unlock understanding of multiple metabolic pathways, and enable accurate determination of gene function. Using *Arabidopsis* and sorghum, this pipeline will identify all metabolites derived from precursors for which isotopically labeled precursors are available and will classify as yet unknown amino acid-derived metabolites and identify genes involved in their production.

### Discovering Innovations in Stress Tolerance through Comparative Gene Regulatory Network Analysis and Cell-Type Specific Expression Maps

- **PI:** José R. Dinneny (Stanford University)
- **Co-Investigators:** Maheshi Dassanayake, Dong-Ha Oh (Louisiana State University); Song Li (Virginia Polytechnic Institute and State University); John Schiefelbein (University of Michigan)

Innovations in gene function allow wild plants to inhabit environments that are commonly stressful to domesticated crops. The goal of this project is to identify such innovations by defining the regulatory and physiological context that these genes function in across several *Brassicaceae* genomes. Using recent advances in

sequencing technologies and machine learning, putative genes associated with extremophyte resilience will be introduced into stress-sensitive species to test whether extremophyte physiological properties can be transferred to naive genomes. The result will be an experimental and data analytics pipeline that will be broadly applicable to the study of gene function at the plant family level, producing a roadmap for improving plant traits.

### Elucidating the Molecular Mechanisms Underlying Drought Resilience in Sorghum

- **PI:** Andrea L. Eveland (Donald Danforth Plant Science Center)
- **Co-Investigators:** Todd C. Mockler (Donald Danforth Plant Science Center), William Duke Pauli (University of Arizona), Brian Dilkes (Purdue University)

The goal of this project is to define and functionally characterize genes and pathways related to drought stress tolerance in sorghum and the molecular mechanisms by which these factors drive phenotypic diversity. Leveraging a field-based phenotyping infrastructure in Maricopa, Arizona, this team will use integrative genomics, high-resolution phenotyping, and molecular and genetic assays with targeted gene editing to characterize and experimentally validate gene function predictions. Identifying the genes and regulatory networks that underlie phenotypic variation in drought tolerance will reveal the genetic basis for the remarkable phenotypic plasticity of sorghum and identify control points for enhancing the productivity of bioenergy crops in marginal environments through precision breeding or engineering.

### Creation of an Acyltransferase Toolbox for Plant Biomass Engineering

- **PI:** Brian Fox (University of Wisconsin-Madison)
- **Co-Investigators:** John Ralph (University of Wisconsin-Madison), Shawn Mansfield (University of British Columbia)

The project's objective is to expand the understanding and utility of plant acyl-CoA ligases and BAHD acyltransferases through a systematic evaluation of all members of these important gene families. An experimentally validated set of up to 30 genes encoding "best-in-class" enzymes for *in planta* screening in *Arabidopsis*, poplar, and sorghum will be generated using *in vitro* biochemistry. Those found to encode desirable functional properties such as growth and cell wall characteristics will be tested in poplar and sorghum. This acyltransferase toolbox will provide a "knock-out/knock-in" list for performing precision engineering of plant metabolism to obtain new and improved properties.

---

## Transforming Our Understanding of Chloroplast-Associated Genes through Comprehensive Characterization of Protein Localizations and Protein-Protein Interactions

- **PI:** Martin C. Jonikas (Princeton University)
- **Co-Investigator:** Danny J. Schnell (Michigan State University)

This project aims to generate a map of protein localizations and protein-protein interactions for chloroplast-associated genes, the energy-producing organelle which is a hallmark of plants. Utilizing the unicellular model green alga *Chlamydomonas reinhardtii* and the biofuel oilseed crop *Camelina sativa*, a searchable online resource for nearly all chloroplast-associated proteins will be generated by leveraging high-throughput protein tagging, microscopy, and affinity purification-mass spectrometry. High-priority localizations and protein-protein interactions will be validated in *Camelina* and used to advance understanding of protein interaction networks that impact yield and stress resistance.

## Constructing the Nitrogen Flux Maps of Plants

- **PI:** Hiroshi Maeda (University of Wisconsin-Madison)
- **Co-Investigators:** Trent Northen, Yasuo Yoshikuni (Lawrence Berkeley National Laboratory); Zoran Nikoloski (University of Potsdam and Max Planck Institute of Molecular Plant Physiology, Germany); Taichi Takasuka (Hokkaido University, Japan)

The objective of this study is to construct plant nitrogen flux maps (NFM) from plant genomes and to determine system-level functionality of a plant nitrogen (N) metabolic network. Biochemical functions of aminotransferase (AT) enzymes will be assessed through high-throughput gene and protein synthesis and enzyme assay platforms, followed by computational protein modeling to predict AT functions from other species and validation by kinetic stable isotope precursor feeding. The resulting NFMs will serve as a framework to quantify N flow through the plant metabolic network, simulate plant metabolic response to N availability, and identify potential targets for improving N use efficiency.

## High-Throughput Determination of a Subcellular Metabolic Network Map of Plants

- **PI:** Seung Y. Rhee (Carnegie Institution for Science)
- **Co-Investigators:** David W. Ehrhardt (Carnegie Institution for Science), Markita del Carpio Landry (University of California Berkeley), Jenny Mortimer (Lawrence Berkeley National Laboratory)

An integrated pipeline combining computational prediction, metabolic network modeling, and high-throughput

experimental testing will be developed using live confocal imaging, nanomaterial-mediated plant transformation, and metabolic network modeling. A high-quality subcellular map of small molecule metabolism will be created in *Sorghum* and *Brachypodium*, which will then be used to generate accurately compartmentalized metabolic network models will be experimentally validated with knock-out gene expression and carbon nanotube-mediated CRISPR technology. This project will create a set of ground-truth data for subcellular locations of metabolic enzymes, yielding important insight into the structure and function of metabolic networks.

## TGCM: (T)rait, (G)ene, and (C)rop Growth (M)odel Directed Targeted Gene Characterization in Sorghum

- **PI:** James C. Schnable (University of Nebraska-Lincoln)
- **Co-Investigators:** Erin Bunting, Addie M. Thompson (Michigan State University); Yufeng Ge, (University of Nebraska); Brandi Sigmon (University of Nebraska-Lincoln); Kaiyu Guan (University of Illinois Urbana-Champaign); Jian Jin (Purdue University); Yumou Qiu (Iowa State University)

The objective of this project is to develop and use tools to improve the efficiency and cost-benefit equation of using reverse genetics to characterize gene function in plants. Machine learning, quantitative genetics, and physiologically informed crop growth models and statistical methods adapted from the BayesCpi approach to genomic prediction will be used to predict key genes, find associations, and identify component traits that contribute significantly to key composite traits such as biomass production. The resulting set of computational, statistical, and genetic tools could be deployed in any plant species to significantly increase throughput for characterization of previously unstudied genes.

## Combining Genome-Wide Association Studies and Expression Quantitative Trait Nucleotide Mapping with Molecular and Genetic Validations to Identify Transcriptional Networks Regulating Drought Tolerance in Populus

- **PI:** Gail Taylor (University of California, Davis)
- **Co-Investigators:** Jin-Gui Chen, Wellington Muchero (Oak Ridge National Laboratory)

The aim of this research is to determine the genetic basis of drought tolerance, a highly complex trait, in bioenergy *Populus*. Natural genetic variation found in two *Populus* populations will be harnessed with available high-throughput phenotyping, computational tools,

---

and extensive sequence resources to identify genetic loci linked to traits and resolve them to the level of cis- and trans-acting eQTL. Findings will be validated for drought tolerance in fast transient assays using protoplasts and in longer-term whole tree experiments with accurate deployment of CRISPR-Cas9 gene edits. This research will deliver new *Populus* genotypes that are better suited and able to tolerate drought environments.

### **Deep Green: Structural and Functional Genomic Characterization of Conserved Unannotated Green Lineage Proteins**

- **PI:** James Umen (Donald Danforth Plant Science Center)
- **Co-Investigators:** Jianlin Cheng (University of Missouri-Columbia); Eric Knoshaug, Vladimir Lunin, Ambarish Nag (National Renewable Energy Laboratory); Ru Zhang (Donald Danforth Plant Science Center)

In this project, a pipeline and associated databases for structural prediction and functional characterization of plant proteins of unknown function (Deep Green proteins) will be established in several model species. To characterize and annotate unknown plant proteins, *in silico* structural predictions and network analyses will be employed to assign structures and predict function to the curated Deep Green candidate protein set. Reverse genetic resources for *Chlamydomonas* will be assembled and curated, then functionally characterized and prioritized. Selected candidates

will be structurally and functionally validated in the reference plants *Arabidopsis* and *Setaria*. The results will aid investigations into the impacts of conserved unknown proteins on photosynthesis, biomass accumulation, and stress response.

### **InferNet: Gene Function Inference by Leveraging Large, Organ-Specific Expression Datasets and Validation of Non-Redundant Regulators**

- **PI:** Kranthi Varala (Purdue University)
- **Co-Investigators:** Ying Li (Purdue University), Karen Hudson (U.S. Department of Agriculture Agricultural Research Service)

This project combines computational and molecular approaches to characterize the role played by key transcription factors (TFs) in regulating seed lipid biosynthesis. Two of the principal challenges in determining gene function are subtle organ- and/or developmental stage-specific phenotypes and functional redundancy among gene family members. A general approach will be developed that addresses both concerns by focusing on regulation of a biological process of interest in an organ-specific manner and by estimating the likelihood of functional redundancy of a given TF. Validated TF regulation knowledge gained will be translated to a bio-fuel crop (e.g., *Camelina sativa*). This project offers a proof-of-concept for an approach that is extensible to any agronomic/biofuel trait of interest.

---

## **Contact and Websites**

### **BER Program Manager**

Catherine M. Ronning  
301.903.9549  
catherine.ronning@science.doe.gov

### **Websites**

**BER Genomic Science program**  
genomicscience.energy.gov

**DOE Office of Biological and Environmental Research**  
energy.gov/science/ber/

**DOE Office of Science**  
energy.gov/science/

**October 2019**