Plan Overview

A Data Management Plan created using DMPTool

DMP ID: https://doi.org/10.48321/D1Q59M

Title: PLANT BREEDING PARTNERSHIP: A Metabolic Modeling-Based Strategy to Accelerate the Citrus Genetic Engineering Process

Creator: James Borneman - **ORCID:** 0000-0003-0661-4970

Affiliation: University of California, Riverside (ucr.edu)

Principal Investigator: Cristal Zuniga, James Borneman, Kranthi Mandadi, Zarsten Zengler

Data Manager: Cristal Zuniga, James Borneman

Funder: United States Department of Agriculture (usda.gov)

Funding opportunity number: A1141

Template: USDA-NIFA: National Institute of Food and Agriculture

Project abstract:

This is a **New Investigator** proposal led by Cristal Zuniga. The **Long-Term Goal** of this proposed project is to improve the economics, productivity, and sustainability of U.S. agriculture by using metabolic models to accelerate the engineering process of all crops. To demonstrate the feasibility of this approach, the **Objective** of this project is to use metabolic models to speed up the growth rate of two specific steps of the citrus engineering process in two commonly grown cultivars. Our team is pioneering the development and application of experimental and computational systems biology approaches to improve growth rates and increase biomass production of phototrophs, **giving this project has a high likelihood of successful**. We expect that this approach will have wide applicability by providing a blueprint for accelerating the engineering process of all crops, **thereby providing considerable benefits to U.S. agriculture**. Since this proposed project is to use metabolic models to accelerate the crop engineering process, **it specifically addresses one of the Program Area Priorities** of the *Plant Breeding for Agricultural Production Program*, which is "the incorporation of modeling (including crop growth models) in breeding." In addition, since our project has the potential to positively impact the engineering of all crops, it therefore also has the potential to address many of the other Program Area Priorities. This project is also a **Partnership with a Minority-Serving Institution** – University of California Riverside – which will provide important opportunities for minority undergraduate students to participate in cutting-edge research.

Start date: 11-01-2021

End date: 10-31-2024

Last modified: 01-23-2024

Copyright information:

PLANT BREEDING PARTNERSHIP: A Metabolic Modeling-Based Strategy to Accelerate the Citrus Genetic Engineering Process

Primary non-digital and digital data generated will come from computational work, lab work, and greenhouse work. Data will be diverse and include hand-written observations, images, videos, nucleotide and protein sequences, genome sequences, genome annotations, metabolic models, metabolomics and transcriptomics data, genetic and phenotypic data, media formulations, and plant growth measurements. Metadata will include collaborating institution, researcher, date, experimental methods, plant growth measurements, media formulations, conditions, locations, and digital file names associated with individual experiments. We will also incorporate the FAIR Framework that can be found at this link – https://www.go-fair.org/fair-principles.

Data formats will be non-proprietary, unencrypted, machine-readable, recognizable by the scientific community, and interoperable among platforms and applications (e.g., TXT, DOC, XML, PDF, CSV, TIFF, and JPEG). In the unlikely event of using a proprietary data format, clear instructions for data access and software source (i.e., software name, version, and company) will be included as a simple text file in the data directory. The format of the metabolic models follows the guidelines presented at BIGG Models (http://bigg.ucsd.edu/), which is the recognized standard in this field. Non-digital data will be digitized by scanning or manual input. Data submitted to public databases (e.g. NCBI), meet all format requirements. A University of California librarians specializing in data services is available to assist with annotation of research data, formatting, and metadata workflows for submission to archiving and for use by the scientific community. We will also incorporate the FAIR Framework that can be found at this link – https://www.go-fair.org/fair-principles.

Labs using digital notebooks and LIMS systems will be backed up on hard drives. Laboratory computers are routinely backed up on hard drives and a cloud system. Lab members' personal computers are backed up monthly on an external hard drive. Datasets for publications will be submitted in Dryad for curation and preservation. The project will also use the University of California Curation Center (UC3) Merritt Repository Service to manage, archive, and share digital content. Merritt provides public access via persistent URLs, tools for long-term data management, and permanent storage options, with built-in contingencies for disaster recovery. All data will be preserved for a minimum of five years after project completion. High-value genome-edited plants will be maintained in greenhouses and tissues sent for long-term storage in the USDA cryopreservation facility at Fort Collins, Colorado. Products including plasmids, nucleic acids, will be retained for at least three years and often longer by storage at -20C and -80C. Plasmids of potential general use (e.g. CRISPR-based genome editing vectors and citrus-specific promoters) will be deposited at Addgene. We will also incorporate the FAIR Framework that can be found at this link – https://www.go-fair.org/fair-principles.

During the grant, data will be deposited in Merritt (as described above) which allows public sharing. Research data will also be cataloged in the Ag Data Commons as required. Final published data will be made publicly available. PDs will deposit papers published without open access in the UC "eScholarship" digital repository. All publications and presentations acknowledge USDA-NIFA support. Datasets on genomes, gene expression, or metabolomics profiling will be available through NCBI or Dryad. Progress and final reports will include a persistent identifier that provides links to the full text. All final data associated with the project will be retained for a minimum of five years after project conclusion or any project publication. If requested, data will be shared with qualified parties, as long as such a request does not compromise intellectual property interests or interfere with a publication. All members of the research team will make presentations at stakeholder events and scientific conferences. We will also incorporate the FAIR Framework that can be found at this link – https://www.go-fair.org/fair-principles.

PD Cristal Zuniga with assistance from Co-PD James Borneman (or if needed their replacements, which would likely be the other project participants Co-PI Mandadi and Co-PI Zengler), will provide oversight of all data management activities and responsibilities. No funds will be needed for data management because will be using all public domain databased, software and/or services. All members of the project's research team with access to data will receive instruction in the Responsible Conduct of Research, which includes proper maintenance of laboratory notebooks. We will also incorporate the FAIR Framework that can be found at this link – https://www.go-fair.org/fair-principles.

Planned Research Outputs

Dataset - "Omics Datasets"

Metabolomics and RNA-Seq datasets of the two citrus engineering steps that we will endeavor to accelerate.

Dataset - "Annotated Citrus Genomes"

Annotated Citrus Genomes

Model representation - "Metabolic Models will by archived at Bigg Models - https://bigg.ucsd.edu/"

Metabolic Models will by archived at Bigg Models - https://bigg.ucsd.edu/

Text - "Plant Growth Media Formations - Provided in Publications & at https://escholarship.org/"

Plant Growth Media Formations - Provided in Publications & at https://escholarship.org/

Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Omics Datasets	Dataset	2024-10-30	Open	NCBI	50 GB	Creative Commons Attribution Non Commercial 4.0 International	None specified	No	No
Annotated Citrus Genomes	Dataset	2024-10-30	Open	NCBI	800 MB	Creative Commons Attribution Non Commercial 4.0 International	None specified	No	No
Metabolic Models will by archived at Bigg Models - ...	Model representation	2024-10-30	Open	None specified	1,000 MB	Creative Commons Attribution Non Commercial 4.0 International	None specified	No	No
Plant Growth Media Formations - Provided in Public ...	Text	2024-10-30	Open	None specified	1 MB	Creative Commons Attribution Non Commercial 4.0 International	None specified	No	No